

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2005, 13:48:02 ; Search time 3265 Seconds  
(without alignments)  
10462.765 Million cell updates/sec

Title: US-10-810-788a-4

Perfect score: 705  
Sequence: 1 aagcttcaattcggga.....ctctgctcattgaataacc 705

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb ba:\*  
2: gb hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_str:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	653.8	92.7	84203	8	AC005106	Genomic s
2	653.8	92.7	103223	8	AC007153	Arabidops
3	73.4	10.4	177522	9	AC009356	Homo sapi
4	71.8	10.2	174080	2	AF217246	Homo sapi
5	71.6	10.2	186175	2	CR318664	Danio rer
6	70.4	10.0	246589	2	CR450749	Danio rer
7	70	9.9	165550	2	AC015830	Homo sapi
8	68	9.6	281723	3	PF292935	Plasmodiu
9	67.8	9.6	18556	9	AY527817	Homo sapi
10	67.8	9.6	174741	9	AC040977	Homo sapi
11	67.4	9.6	225328	2	CR847887	Danio rer
12	67.2	9.5	165797	2	CR847851	Danio rer
13	67	9.5	125623	3	AC115599	Dictyoste
14	66.8	9.5	183394	2	CR753884	Danio rer
15	66.6	9.4	136762	2	CR753796	Danio rer
16	66.6	9.4	253755	2	CR812465	Danio rer
17	66.2	9.4	209491	2	CR392365	Danio rer
18	66	9.4	169448	2	CR385039	Danio rer
19	66	9.4	273275	3	AE014828	Plasmodiu

20	65.6	9.3	273275	3	AE014828	Plasmodiu
21	65.4	9.3	177444	2	AC146461	Pan trogl
22	65.4	9.3	181430	2	CR407555	Danio rer
23	65.4	9.3	183344	9	AC147112	Pan trogl
24	65	9.2	185342	2	CR387989	Danio rer
25	64.8	9.2	149831	2	CR847943	Danio rer
26	64.4	9.1	183648	3	AC117081	Dictyoste
27	64.2	9.1	18056	6	AX559046	Sequence
28	64.2	9.1	174574	2	CR759746	Danio rer
29	64	9.1	191689	2	BX957320	Danio rer
30	64	9.1	211643	5	BX530036	Zebrafish
31	63.8	9.0	148486	2	CR753871	Danio rer
32	63.8	9.0	333321	3	AC116986	Dictyoste
33	63.6	9.0	250823	3	AE014821	Plasmodiu
34	63.4	9.0	611	6	AX187051	Sequence
35	63.2	9.0	253151	3	AE014842	Plasmodiu
36	63	8.9	110000	2	PFMAL13_24	Continuatio
37	63	8.9	146570	3	AC117076	Dictyoste
38	63	8.9	182870	3	AC116960	Dictyoste
39	62.8	8.9	250022	3	AE014824	Plasmodiu
40	62.8	8.9	250793	2	CR388030	Danio rer
41	62.6	8.9	1453	3	AF483206	Planococc
42	62.6	8.9	8056	6	AX559046	Sequence
43	62.6	8.9	85916	3	AC117080	Dictyoste
44	62.6	8.9	113880	3	PFMAL3P4	Plasmodiu
45	62.6	8.9	168069	2	CR812470	Danio rer

## ALIGNMENTS

RESULT 1	AC005106	84203 bp	DNA	linear	PLN 28-JUN-2000
LOCUS	AC005106	Genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I, complete sequence.			
DEFINITION	AC005106	Genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I			
ACCESSION	AC005106	GI:6587719			
VERSION	AC005106.2				
KEYWORDS	HTG				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1 (bases 1 to 84203)				
AUTHORS	Johnson-Hopson, C., Brooks, S., Buehler, E., Chao, Q., Khan, S., Kim, C., Shim, P., Altati, H., Bei, Q., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howang, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharasy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaver, A., Tortum, M., Vaysberg, M., Yu, G., Federspiel, N.A., Theologis, A. and Ecker, J.R.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 84203)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submision				
JOURNAL	Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	3 (bases 1 to 84203)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submision				
JOURNAL	Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	4 (bases 1 to 84203)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submision				
JOURNAL	Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				

[illegible]



```
gene
CDS
    complement(29404..40201)
    /gene="F3F20.1"
    /note="Highly similar to putative callose synthase catalytic subunit; Highly similar to cotton putative callose synthase catalytic subunit, gb|A025952"
    /codon_start=1
    /protein_id="A020609.1"
    /db_xref="GI:4836907"
    /translation="MAQRRPDPPEPPORILRTQTVSGEAMLDSEVPSSVLEIAP
    IIRVANEASNPVAVLCRFYAFKARHLDPSGSGROKFTALLOLERENETTL
    AROKSDAREMOSFOHYKKYKYOALNLNADKADRAOLTKRQYOTAVLEVKAVNQT
    EDEVDADENVLIDYDILITQIYVYPIILLDPSNOAMLPETIOAAVAALNRG
    LPTAGHAKKLLDEDLIDMLOSFGKQVNLNREHLLILANVHLKOPKPDQOKL
    DDRALITVKKLFRNYKKWKCYLGRSSLMPTIQEVOQKLLYGLVLLWGEAN
    LFMPCLCYIYHHLGMLAGSVPMTGHNKPAQGEDEAFLOKVTVPITYOTISKEA
    KSGRGSKSHVWRNYDDLNIEFWSIRCFRLGPMRADPFCOTAELELSLESIS
    NSGDRMGKVNFEIRSPFHWIIFSRPRLMSFYILCLOMIVITANGSGELSAPOD
    ELKVLVPTTAALKAQAVLDIALSWKARSHSLVYKRYWKVGAAYVYVMAVT
    VAYSKNASGFSQTIKMFEGSHNSPSLVAIILYLSBNMLSLLEPPIRRLVE
    RSDYKIMLMWMMWSQSPSPSSMOIKPLVPTDINRHHVSVSMHFPFPAKNINLV
    VIALVSPVLLVEMDQIYVAIVSTLVGLNGAFRLGEIRLGMKLSRSPFQAFN
    DCLVDPNSDDTCKKRAFATPSKFPOLPSKQDEARPAQMMNKIISFREEDLSD
    REMELLILVYMSDPLDILRMPRLIASKIPIALDMAKDSNGKRELKRIKRLVDSMT
    CAVRECYASFQKILNYLVGERGQYINDIFSKIDHIKRTILITLANSALPDLQ
    FRLIYELNREEDKOQIVILNMLELITRDMEEVPSAIVSNFOSPLKRLK
    GKKKQIKRLHLITVRESAMDVPSNLBARRLIFPNSNLFMDPPAPKIRNMLSPFL
    TTYFSEDLVFSIFGLEQNEDEVSILFYLOKIPDEMTNLFERKVCNGBEELIARLD
    BEELRLMASVRCQTLTKTVGRMYKRALELQAFDMADKDELKGLYKLELTSEBAS
    GSGSLMAOQCALADMKFTPVVSCQOYSIHKSGDRAOILRLMTTVYSIRAVYDE
    VQOTKESYKGTBEKIYSALYKAAQPTKPMBSSEVQTLDOIYIKLIPRAIILBEG
    KENONHAIIFTRBGLQITIDMKNQYMEAFKRNRLQEFLEKHGVCPCPTLLAGE
    HIFGVSLSLWFMNSQENSFVTIQQGVLASPLKAFPHYGHPIIDRLPHLRGFSNT
    LEBGVNTHIEYIIOVGKRDVGLNQISMPFAKILANGGEQTLISDLIRLRFPFML
    SCFTTIGFYFSTMLTVLVYVYFLYGLRVLVLSGLEGLSSORAFNNKRLPEALASO
    SFVOIGFLMALPMMEIIGLERPHNALIRFVLMOLASVFPFPOCTGTHYGRTLF
    HGALEYRGTRGRTGVYVNAKFAENRYRYSNHPKGLIELMILLVIOIFQOSIRGVTY
    ILLTVSIPMVVWTWLFAPLFNPSGFEMQKIVDDWDMNKKIYNRGIOGVPPKPSX
    WMEKLEHLRSGVGRITLIEILALRFLFOYGLVYHLSTFKGNQSPFWYGVASFWVI
    LEILLIVKGLRGVRRTSTNFOQLFRILKGLVFLVAILITFLALPILITIKDLFCM
    LAEPFGMGLMLIAQCKPLIQOLGIGWSSRTIARGYELVGMGLLFTVPAPLAMPFV
    SFEQTRMLFNQASRGLQISRLIGGORKRSRKNKE"
    complement(44167..45812)
    /gene="F3F20.2"
    complement(join(44167..44362,44438..44542,44760..44528,
    45331..45446,45524..45593,45680..45812))
    /gene="F3F20.2"
    /note="Hypothetical protein"
    /codon_start=1
    /protein_id="A030611.1"
    /db_xref="GI:4836909"
    /translation="MESNGGGEVRRVNLVYFLRSRGVHDHHLRLVHLSNGVFLR
    DYKMLADARGAMPDAFSMSCRRYKNGVYMODLDDLLIPISNEVNLKGSLELL
    SSKEDYVNEKKAWYTRNGCIDAEKLOKLTSEKIOKESVPSCSSTATTSYV
    ESTTNEBGFVLKKQDPKTVSGGDSGTEGSGNDVESGSPVSTSSSTIKNSY
    SSVRASHVLRNLNMGCGDLTNDVAVLPLNKSRSAGAPMEBERRQYOHQNNARFS
    BEAWSGIKMKETIEPCPKVAPSKPMALPSCGCLPPEKMKHSGKLCRGKNSA
    NNDLMTSNNTVPRQORCNIPGNPLGHRVLTTLTKE"
    complement(47118..50318)
    /gene="F3F20.3"
    complement(join(47118..48467,48761..49149,49186..49302,
    49523..49979,50064..50318))
```

```
gene
CDS
    /gene="F3F20.3"
    /note="similar to Na/H antiporter proteins; N-terminal half of protein is similar to Na/H antiporter proteins"
    /codon_start=1
    /protein_id="A030610.1"
    /db_xref="GI:4836908"
    /translation="MSGAPLVNTNPNVYIEBSRFEKICYDOSLLFEKRGKMSG
    STLASSLPFIIQLFVANSYRVLVLPVLPVPAOILCGLSPSVLQKTRPI
    AHVPRFPMVLETFANLALVNIPLFGMDLRMRITELKVIILFGLVLAFLVG
    AFVLYPFGSHDCKIISGCVFMSVALACNPDIAILDLIDKDMRSTACCAIV
    TDLCTWVLLVFGPASFSGSKGTNNKMPVITTAIVLCIFIRGDAWIFAKYKA
    GGVVLCGLITDPCGHSITGAPLPGLSIHDHIRMIEKILDELISGLMPLFYIC
    GLRADIGMLOTTDKRMVNVICSPFLVXITVTVITSLTMHLMRDAFALGALMTKG
    TLSLVNAGRDPTKALDSPYTHMTALLVMSLVPEPLAFAPKPKLAAHYKRTVQ
    KIGETELRVLCVHLPNVSGITNLIQVSNATKOSPISVFAHLVLELGRITASLLI
    NMDECKPKNFDRVVAESDOJLAETFEAMEVNNDAMTVOTITVAVSATMHEDICVLA
    EKRVCFTILPYHKHLTPDGRGEGNSHAEINOVNLSHAPCSVGLVLDVGRGMARSE
    SPRGSEKREVMALPYGPDPRALASVARMQCHYIKLVVRFVGRALISGGVA
    AERYEKQVDECTYFNRKTNMDSVKTIEKYVNDQOTITATIREMEDNSLDIVV
    GRGVNSDSPVTAGLNDSSPELGTIDGLASNFTMHASVAVIQOYSATKROAAVTA
    AATTVMGAVTGAQGNLLESAGDAMKTRDAHEPFMKSYEDEDEEDHOGIHR"
    51395..53210
    /gene="F3F20.4"
    join(51395..52184,52258..53210)
    /gene="F3F20.4"
    /note="Similar to hexosaminidase; Similar to hexosaminidase, gi|1170249, gi|1076942 and others"
    /codon_start=1
    /protein_id="A030612.1"
    /db_xref="GI:4836910"
    /translation="MLTSLKRFVILIPILFTTLSPFLSIALPINIKRPFSLSPQ
    HAAIALSPFTLAPHEQYLSVYRKNVLIISENSPLISYPVKLMKRYTLRNLT
    VYDFSLPHNGDEYSKLSIPGSEFAHLIASAMVAMKLETSQIMTDSGLCP
    VYDIPDSLPFGHGVLDLSTSNRYGVDDIMETIKMSAMKLNFMHMTDSQFPLV
    LSEBPLAKSGIPMVVTPPDQSLIYVQYEGHGRVLPETIDPSTGWSGAEI
    VYCANMFPAKSKMWERLASRPGQALNPLSKPYKEVYVNIQDI VNOPSPSRFEG
    GBEVLPFGKTRPALNSFLSSGTTLSQLEKLTINSTLPYIVSQNRVTVMEVDLDAQ
    IADSPVLPKEHTIQLTNNGENPNTRIYVAQYRVIVSSSEYVLTDCGGHFGANDS
    IYQDSGSGSCWCAPEKTMQSIYNYDIADDLNBEERKVLVGGSEVALMSQASDVTL
    RYALPMSALAESLNSGRNDRBVRKCGEAYDLNIRYVYKRGIGABPIQPLMCLKN
    PGMCTVHGAALDOQ"
    53685..55189
    /gene="F3F20.5"
    53685..55199
    /gene="F3F20.5"
    /note="Hypothetical protein"
    /codon_start=1
    /protein_id="A030626.1"
    /db_xref="GI:4836924"
    /translation="MSVVRMPRLVTPSLISQILKKOKNPTALKLFBAKERPPSYGH
    NSVAVATMIDIIGKSNRVLEMKVIERMKEDSCCKDSVFAVIRTFBSRGRLEDAIS
```

```
Query Match 92.7%; Score 653.8; DB 8; Length 103223;
Best Local Similarity 99.1%; Pred. No. 6.7e-94;
Matches 699; Conservative 0; Mismatches 2; Indels 4; Gaps 4;
```

```
QY 1 AAGCTTTAAATTTGGGAAAAATTTAATATATGATTAATATG-CAATATAAAGTAG 59
Db 20166 AAGCTTTAAATTTGGGAAAAATTTAATATATGATTAATATGCAATATAAAGTAG 20225
QY 60 ATGAGAGTTTTTTTAAATTTTATTAATAAATAA-TTTTTTGATGAAATGTTTTTA 118
Db 20226 ATGAGAGTTTTTTTAAATTTTATTAATAAATAAATTTTTCGACGAATGTTTTTA 20285
QY 119 GATAAAATTTTGACATTAACCAAGTATATTTTCCGACGAATTTGAGTCTACGAGGAG 178
Db 20286 GATAAAATTTTGACATTAACCAATATATATTTTCCGACGAATGAGTCTACGAGGAG 20345
QY 179 TCTGTCTTAAACCTCGTAAAGTTTGGCTTACCAACCAACCAATCCCAAAAGGTAA 238
Db 20346 TCTGTCTTAAAC-CTCGTAAAGTTTGGCTTACCAACCAACCAATCCCAAAAGGTAA 20404
QY 239 CGACCATACCGGTATATAATTTCTAACCAGTTTATAGTTTACATTAATCATTTACTAA 298
```

```

Db      20405 CGA-CATTAACGGTTAAATTAATTCATACCCGTTTATTAAGTTTACATAAATCATTTACTAA 20463
Qy      299 TCGCGCGTAATTAAGTTTATGAAANGGTTATTTTGTAGGACAGCTAAATTTATTA 358
Db      20464 TCCGCGGTATTAATTAAGTTTATGAAANGGTTATTTTGTAGGACAGCTAAATTTATTA 20523
Qy      359 ATTTTGTAGTACTGTTTCTTTTGGGTTCAACTAGTACTTTTCTTGTGACATC 418
Db      20524 ATTTTGTAGTACTGTTTCTTTTGGGTTCAACTAGTACTTTTCTTGTGACATC 20583
Qy      419 AAAATTATTTGTAGACAGATGCTCCATATATAGATGTGAATGAAATGAAATTTAGTAA 478
Db      20584 AAAATTATTTGTAGACAGATGCTCCATATATAGATGTGAATGAAATGAAATTTAGTAA 20643
Qy      479 TAAATTAATTAATTAAGTGAACAAAAAATTAAGTGAACAAAAAGTGAACAAAAAG 538
Db      20644 TAAATTAATTAATTAAGTGAACAAAAAATTAAGTGAACAAAAAGTGAACAAAAAG 20703
Qy      539 CTATTATGCTTGGCAAGATCCATTAAGTTATTTCACCAAAAAGAAAGAGTACTTGCC 598
Db      20704 CTATTATGCTTGGCAAGATCCATTAAGTTATTTCACCAAAAAGAAAGAGTACTTGCC 20763
Qy      599 TTGCTCTCTCTCTCTTTTAAATGTAAGTGAAGTGAAGATGAGACTCAATTTCTAGT 658
Db      20764 TTGCTCTCTCTCTCTTTTAAATGTAAGTGAAGTGAAGATGAGACTCAATTTCTAGT 20823
Qy      659 CATCAAAACATATTTGGCGCTTATTTCTGTGCTCACTGGAATA 703
Db      20824 CATCAAAACATATTTGGCGCTTATTTCTGTGCTCACTGGAATA 20868

```

```

RESULT 3
AC093566 177522 bp DNA linear PRI 29-NOV-2001
LOCUS Homo sapiens chromosome 8, clone CTD-2017M22, complete sequence.
AC093566
VERSION AC093566.3 GI:16905291
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 177522)
AUTHORS Birren, B., Linton, L., Nuebaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone CTD-2017M22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177522)
AUTHORS Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
Brown, A., Camarata, J., Campiano, A., Chang, J., Chararo, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talama, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trisillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

TITLE  
JOURNAL  
Submitted (03-SEP-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

REFERENCE 3 (bases 1 to 177522)
AUTHORS Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
Brown, A., Camarata, J., Campiano, A., Chang, J., Chararo, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talama, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trisillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE
JOURNAL
COMMENT Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 13, 2001 this sequence version replaced gi:16041411.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12730
Center clone name: 2017_M_22
-----
FEATURES
source
Location/Qualifiers
1. 177522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="CTD-2017M22"
/clone_lib="CTD Human BAC"
106..781
repeat_region
/rpt_family="MER21B"
794..1024
repeat_region
/rpt_family="WIR"
1654..1695
repeat_region
/rpt_family="L2"
2458..3178
repeat_region
/rpt_family="L1MA7"
3180..3323
repeat_region
/rpt_family="AT_rich"
complement(3335..3986)
/rpt_family="L1MCD"
complement(4057..4520)
/rpt_family="L1MA3"
4521..5565
repeat_region
/rpt_family="THR1_int"
complement(5560..7819)
/rpt_family="L1M1"
complement(7834..8206)
/rpt_family="L1MCD"
complement(8207..8511)
/rpt_family="AluSx"
complement(8512..8628)
/rpt_family="L1MCD"
8629..8683
repeat_region

```

repeat\_region /rpt\_family="(TG)n"  
complement(8684. .10137)  
/rpt\_family="L1MCB"  
repeat\_region 10197. .10266  
/rpt\_family="L1MCA"  
repeat\_region 11656. .11678  
/rpt\_family="(TTTG)n"  
complement(12114. .12294)  
/rpt\_family="MIR"  
repeat\_region 12754. .12886  
/rpt\_family="L2"  
repeat\_region 13171. .13530  
/rpt\_family="MLT2P"  
repeat\_region 13531. .13564  
/rpt\_family="(TG)n"  
13565. .13608  
/rpt\_family="MLT2P"  
complement(14319. .14678)  
/rpt\_family="AluSx"  
repeat\_region 15147. .15224  
/rpt\_family="MIR3"  
repeat\_region 15706. .16139  
/rpt\_family="LTR16A"  
repeat\_region 16248. .17519  
/rpt\_family="HERV16"  
repeat\_region 17586. .17815  
/rpt\_family="LTR16A"  
repeat\_region 17995. .18020  
/rpt\_family="AT\_rich"  
repeat\_region 18473. .18715  
/rpt\_family="MIR"  
repeat\_region complement(18761. .18886)  
/rpt\_family="MIR"  
repeat\_region 19366. .19416  
/rpt\_family="(CGGA)n"  
repeat\_region complement(1951. .19672)  
/rpt\_family="MER94"  
repeat\_region 20354. .20660  
/rpt\_family="AluSx"  
complement(20741. .20745)  
/note="<30 qual SINGL region"  
complement(20772. .20776)  
/note="<30 qual SINGL region"  
complement(20781. .20785)  
/note="<30 qual SINGL region"  
repeat\_region 21910. .22230  
/rpt\_family="THE1B"  
repeat\_region 22271. .22297  
/rpt\_family="(CA)n"  
repeat\_region 23141. .23161  
/rpt\_family="AT\_rich"  
repeat\_region 24005. .24077  
/rpt\_family="MLT1T"  
complement(24534. .24673)  
/rpt\_family="L1MA10"  
repeat\_region 24692. .24736  
/rpt\_family="T-rich"  
complement(24744. .25006)  
/rpt\_family="L1MB1"  
repeat\_region 25663. .25687  
/rpt\_family="AT\_rich"  
repeat\_region 26020. .26121  
/rpt\_family="L2"  
repeat\_region 26133. .26375  
/rpt\_family="L2"  
repeat\_region 26376. .26738  
/rpt\_family="THE1B"  
repeat\_region 26739. .27286  
/rpt\_family="L2"  
repeat\_region 27360. .27737  
/rpt\_family="L1MCB"  
repeat\_region 27799. .27899  
/rpt\_family="L1MCB"

repeat\_region 27900. .28212  
/rpt\_family="L1MC2"  
repeat\_region complement(30049. .30160)  
/rpt\_family="L2"  
repeat\_region complement(30217. .30339)  
/rpt\_family="MIR"  
repeat\_region complement(30368. .30563)  
/rpt\_family="L2"  
repeat\_region 31519. .31672  
/rpt\_family="MIR"  
repeat\_region 32015. .32064  
/rpt\_family="L2"  
repeat\_region 33826. .33862  
/rpt\_family="AT\_rich"  
repeat\_region 33914. .33969

Query Match 10.4%; Score 73.4; DB 9; Length 177522;  
Best Local Similarity 46.5%; Pred. No. 0.0071;  
Matches 236; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 5 TTTTAAATTTGGGAAATTTAATAATATATGATTTAAATGCGAATAAAGTTACATGAG 64  
DB 86760 TGTATATATATATATATATATATGATGATGTTTATATGATATATATATATATATG 86819  
QY 65 AGTTTATTAATTTTATATATAAATATTTTTCATGATGATGTTTATGATGATGTTTATGATGATG 124  
DB 86820 ATATGTTTAT 86879  
QY 125 ATTTGACATTAACAAGATTTATTTCCGCCAGCAAGTTAGTCTACGAGAGATGCTGT 184  
DB 86880 TGTTTAT 86939  
QY 185 CTTTAAACCTCGTAAAGTTTGGTCTTACCCCAACCAATCCCAAGTAAAGACCA 244  
DB 86940 ATATCTATATGTTTAT 86999  
QY 245 TACCGTAT 304  
DB 87000 AAT 87059  
QY 305 TGTATTAAGTTTATGAATGTGTTATTTTGTAGTCCATGCAATTTATTAATTTT 364  
DB 87060 TATATGTAT 87119  
QY 365 TAGTACTGTTTCTTTTGGGTTCACATGACTACTTTTCTTGCATCAAAAT 424  
DB 87120 TTGTATATGTTTACATGATGATATATATATATATATATATATATATATATATATATAT 87179  
QY 425 ATTGTAGACAGTGTCCAT 484  
DB 87180 AT 87239  
QY 485 AATATGAAGTGACAAAATAATTA 511  
DB 87240 CATGTAT 87286

RESULT 4  
AF217246/c 174080 bp DNA linear HTG 26-JUL-2002  
LOCUS Homo sapiens chromosome 8 clone CTD-2017M22 map 8q24, WORKING DRAFT  
DEFINITION  
SEQUENCE, 20 unordered pieces.  
ACCESSION AF217246  
AF217246.5 GI:14327861  
VERSION HTG: HTGS PHASE1; HTGS\_DRAFT; HTGS\_CANCELLED.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 174080)  
Schlhubel M.B., Baumgart C., Blechschmidt K., Detke M., Jahn N.,  
Lehmann R., Menzel U., Polley A., Reichwald K., Schudy A.,  
Siddiqui R., Taudien S., Wen G., Rosenthal A. and Platzer M.



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

**COMMENT**

Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 186175)  
McLay, K.  
Direct Submission  
Submitted (04-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk  
On Mar 7, 2004 this sequence version replaced gi:45126687.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

[illegible]

## FEATURES

```
misc feature
1..18331
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-59L22"
/clone_lib="CHORI-211"
1..18331
```

```

/note="assembly_fragment:01454
fragment_chain:1
clone_end:5p6
vector_side:left"
18422. .35023
/note="assembly_fragment:00754
fragment_chain:1"
35124. .39045
/note="assembly_fragment:00149
fragment_chain:1"
39146. .52645
/note="assembly_fragment:00580
fragment_chain:1"
52746. .59267
/note="assembly_fragment:00391
fragment_chain:1"
59368. .62964
/note="assembly_fragment:00039
fragment_chain:1"
63065. .70357
/note="assembly_fragment:00484
fragment_chain:1"
70458. .73685
/note="assembly_fragment:00105
fragment_chain:1"
73786. .92113
/note="assembly_fragment:01198
fragment_chain:1"
92214. .94992
/note="assembly_fragment:00071
fragment_chain:2"
95093. .99211
/note="assembly_fragment:00253
fragment_chain:2"
99312. .105947
/note="assembly_fragment:00314
fragment_chain:2"
106048. .136527
/note="assembly_fragment:02182
fragment_chain:2"
136628. .166850
/note="assembly_fragment:01736
fragment_chain:2"
166951. .183315
/note="assembly_fragment:00962
fragment_chain:2"
183416. .186175
/note="assembly_fragment:00199.0"

```

ORIGIN

Query Match	10.2%	Score	71.6	DB	2	Length	186175
Best Local Similarity	38.8%	Pred. No.	0.013				
Matches	209	Conservative	0	Mismatches	329	Indels	0
						Gaps	0

[illegible]



QY	306	GTAAATTAGTTTATGAAAGTGGTATTTTGTAGACGACAGTGAATTTATTAATTTT	365
Db	92069	TTAATATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	92010
QY	366	AGTACTGTCTTTTCTTTTGGGTCACAGTACTTTTCTTTGACATCAAAATTA	425
Db	92009	TTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTTATTTATTTATTT	91950
QY	426	TTGTAGACGACTGTCATATTTATGATGCGAAATGAAATGAAATTTAGTAAATTA	485
Db	91949	TAT	91890
QY	486	ATATGAAAGTGAACAAAAAATTTAGTGAAGAAAGTGAACGTAAGAAAGCTTAT	543
Db	91889	ATAAT	91832
RESULT 6			
CR450749			
LOCUS			
DEFINITION	CR450749	246589 bp	DNA linear HTG 24-MAY-2004
ACCESSION	CR450749		
VERSION	CR450749.1	GI:47604407	
KEYWORDS	HTG; HTGS PHASE.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;		
	Cypriniformes; Cyprinidae; Danio.		
	1 (bases 1 to 246589)		
	Sims, S.		
REFERENCE	Submitted (21-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,		
AUTHORS	CambridgeShire, CB10 1SA, UK. E-mail enquiries:		
JOURNAL	zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
COMMENT	----- Genome Center		
	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: zfish-help@sanger.ac.uk		
	----- Project Information		
	Center project name: ZK206		
	----- Summary Statistics		
	Assembly program: XGAP4; version 4.5		
	Chemistry: Dye-terminator; 100% of reads		
	Consensus quality: 243112 bases at least Q40		
	Consensus quality: 243477 bases at least Q30		
	Consensus quality: 243996 bases at least Q20		
	Insert size: 245589; sum-of-ctigs		
	Insert size: 241608; 2.3% error; agarose-fp		
	Quality coverage: 7.67x in Q20 bases; sum-of-ctigs Quality		
	Coverage: 8.07x in Q20 bases; agarose-fp		
	-----		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 11 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	1 28228: contig of 28228 bp in length		
	* 28328: gap of 100 bp		
	* 28329 39447: contig of 11119 bp in length		
	* 39448 39547: gap of 100 bp		
	* 39548 48874: contig of 9337 bp in length		
	* 48875 48974: gap of 100 bp		
	* 48975 49511: contig of 48577 bp in length		
	* 49512 97651: gap of 100 bp		
	* 97652 184544: contig of 86693 bp in length		
	* 184545 184644: gap of 100 bp		

Query	Best Local Similarity	Score	DB	Length	246589;
Matches 186;	Conservative	0;	Mismatches 383;	Indels 0;	Gaps 0;
Db	184659	AAATTNNATATNNANNTAANAATTATTTTNNNNNAATTNNNNNTTNNANNNNNN	184718		
Qy	62	GAGAGTTTTTTAAATTTTATTTAATAAATAATTTTTCATGAAATGTTTTTAAGAT	121		
Db	184719	ANNNNNTTAAANNAANTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	184778		
Qy	122	AAAATTTTACATACCAAGTATATTTTCGCCACGAATTGAGCTACGAGAGATCTCC	181		
Db	184779	NNNNNNNNNNNTNN	184838		
Qy	182	TGCTTTAAACCTCGTAAAGTTTGGCTTTACCCACCAATACCCCAAGGTAAACGA	241		
Db	184839	NN	184898		
Qy	242	CCATACCGGTATATATTTCTACACGGTTTATAGATTACATATATCATTTACTATCC	301		







/translation="MGPPVUKITSMLSHSEFFNYHISLPILIOYKLNKSSNN  
KIIVDEIISFLIKENIHNDKSKKGYVEGYPNNNLQAYSLKLPKSHVFLYADE  
EYIKYKEENDIAIFSYOKDYDINEHININIKPLKQDVLISYINISDMLTL  
GNKKVNLHDFDQMLIDHVKNEVSKNDDEMSTNGDI"  
join(17997, .18035,18047, .18112,18218, .18278,18587, .18845,  
18040, .19172)  
/gene="PF11555w"  
join(17997, .18035,18047, .18112,18218, .18278,18587, .18845,  
19040, .19172)  
/gene="PF11555w"  
/note="5 probable transmembrane helices predicted for  
PF11555w by TMHMM2.0 at aa 28-45, 60-78, 85-107, 127-149  
and 162-181"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD5197.1"  
/db\_xref="GI:23505217"  
/db\_xref="UniProt/TrEMBL:O812J6"  
/translation="MTLKKIYNGHFRVRVKKKKKKKKKKYIYIKVDGCSIHVS  
LSSEKADKSGFCRGYIVNNILYAGMFLYIIIMSCQKRFYIVYHGFYLSIF  
YLFELHFRSMCMIPASYEKENVLSFCSFLDIFLCKHIIICVFIIFNSTEICKLN  
VNIILHIYIYIYICATFVOKNK"  
complement(join(19209, .19255,19418, .19474,19586, .22846,  
23133, .23511))  
/gene="PF11560c"  
complement(join(19209, .19255,19418, .19474,19586, .22846,  
23133, .23511))  
/gene="PF11560c"

Query Match 9.6%; Score 68; DB 3; Length 281723;  
Best Local Similarity 47.0%; Pred. No. 0.04;  
Matches 247; Conservative 0; Mismatches 275; Indels 4; Gaps 1;

QY 1 AAGCTTTAAATGGGAAATTAATATATATATTAATGAAGCAATTAAGTTAG 60  
DB 168592 AAAATTAGAAAAATCAATTAAAAAGGAAAAAATTTGA 168533  
QY 61 TGAGAGTTTTTTTAAATTTTAAATTTTTCATGCAATGTTTAAAG 120  
DB 168532 TTGTATTTCTTTAATATATATATATATTTTTCATATTAATTTCCAT 168473  
QY 121 TAAATTTTGACATCAAGATTAATTTTCCGCAAGATGAGTACGAGAGATGTC 180  
DB 168472 TATATATCTATATATGAAAAATTAATTAATTTTATATATATCAAGTA 168413  
QY 181 CTGTCTTAAACCTGTAAGTTTGTCTTACCAACCAATACCAACCAAGTAACG 240  
DB 168412 ATATATATATTAATTAATGCAAGATTTCTTAATAGAAAAAAGAA 168353  
QY 241 ACCATACCGTTATATATATTTTACCGTTTAAAGTTATCATTAATCTACTATC 300  
DB 168352 AAAATATATATATATTAATTAATTTTAAAGCAATATATATATTAATTAAC 168297  
QY 301 CGCGTGTAAATTAAGTTTAAATGATGTTTGTAGTCAAGTGAATTTATTAAT 360  
DB 168296 CTTTAAATTAATATATATATATATATATATATATATATTAATTAATTAAG 168237  
QY 361 TTTTATAGTCTGTTTCTTTTGGGTTCACTGTTACTTTTCTTTGACATCAA 420  
DB 168236 TATATTTTAAATTTTATTTTATTTTCTCATATAGAAATATTTTCATTGAAC 168177  
QY 421 AATTATTTAGACGATGTCATATATATGATGAAATGAATTAATTTGATTAAT 480  
DB 168176 ATTTATATGCAATATTTGATATATTTAAATAGGGTGAATTAATTAATTAATTTGTAATA 168117  
QY 481 AATTAATATGAAGTGAACAAAAAATTAAGTGAAGAAAGTAAC 526  
DB 168116 TATATATATATATATATATATTAATTAATTAATTAATTAATTAAT 168071

RESULT 9  
AY527817/c 18556 bp DNA linear PRI 31-JAN-2004  
LOCUS

DEFINITION Homo sapiens arachidonate 12-lipoxygenase (ALOX12) gene, complete cds.  
ACCESSION AY527817  
VERSION AY527817.1 GI:41324131  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 18556)  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
TITLE Rieder, M.J., Daniels, R.L., da Ponte, S.H., Hastings, N.C.,  
JOURNAL Ahearn, M.O., Rajkumar, N., Yi, Q. and Nickerson, D.A.  
COMMENT Submitted (15-JAN-2004) Genome Sciences, University of Washington,  
1705 NE Pacific, Seattle, WA 98195, USA  
To cite this work please use: SeattleSNPs, NHLBI HL66682 Program  
for Genomic Applications, UW-FHCRC, Seattle, WA (URL:  
http://pga.gs.washington.edu).  
FEATURES  
source  
1. .18556  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
repeat\_region  
167, .237  
/rpt\_family="MIR"  
variation  
386 \_type=dispersed  
/frequency="0.56"  
/replace="a"  
864  
/frequency="0.01"  
/replace="c"  
1138  
/frequency="0.89"  
/replace="a"  
1158  
/frequency="0.42"  
/replace="c"  
1440  
/frequency="0.13"  
/replace="a"  
1586  
/frequency="0.01"  
/replace="a"  
1649  
/frequency="0.01"  
/replace="a"  
1661  
/frequency="0.42"  
/replace="c"  
1764  
/frequency="0.29"  
/replace="a"  
1947, .16598  
/gene="ALOX12"  
join(1947, .2114,2688, .2889,4372, .4453,4578, .4700,  
4816, .4919,5169, .5329,6199, .6342,7465, .7674,11122, .11208,  
11723, .11892,12351, .12472,15612, .15712,15821, .15991,  
16109, .16598)  
/gene="ALOX12"  
/product="arachidonate 12-lipoxygenase"  
join(1980, .2114,2688, .2889,4372, .4453,4578, .4700,  
4816, .4919,5169, .5329,6199, .6342,7465, .7674,11122, .11208,  
11723, .11892,12351, .12472,15612, .15712,15821, .15991,  
16109, .16288)  
/gene="ALOX12"  
/codon\_start=1  
/product="arachidonate 12-lipoxygenase"  
/protein\_id="AAS00094.1"  
/db\_xref="GI:41324132"  
translation="MGRRIRVATGAMLFSSGYNRVQMLVGTREBAIEIQLPARG  
EEEDPDVAEDLGLQFRLKXKHMVLVDAMFCDRITVGGGACAEVAFPCYRWVG  
EDILSPEGTARLPQDNALDMFQKREKELKDRQOIYCWATWKEGFLTIADRKDIL



Qy	268	GGTTATTAACGTTATACATAAATCAATTCATTCATACGCGTGAATTAAGTTATTAAGAAATG	327
Db	10173	AAATTTAAATTTTAATTTAATTAATTAATTAATTAATTAATTAAGTTAAATTAATTAATTAATTA	10114
Qy	328	GGTTATTTTGTAGTCACGTGAATTTATTAATTTTTTGTACTGTTGTTTTCTTTTTGG	387
Db	10113	AAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATG	10054
Qy	388	GTTCAACTAGTACTCTTTTCCCTTGCATCAAAATATATGTAGACGAGTGCATATTA	447
Db	10053	TTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	9994
Qy	448	TAGATGCTGAATGAATGAATATTGAGTAAATAATTAATTAATTAATTAATTAATTAATTAATTA	507
Db	9993	TAAATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTA	9934
Qy	508	ATTAGTGAAGAAGTAACTGGAAGAAGAAAGCCTATTGCGTTGGCAAGATTCAT	562
Db	9933	AAATTTTAAATTTTAAATTTTAAATTTTAAATTTTCTTAGCTATTAATTAATTAATTAATTAATTA	9879

RESULT	10		
AC040977/c			
LOCUS	AC040977	174741 bp	linear
DEFINITION	Homo sapiens	chromosome 17, clone Rpl1-589p10, complete sequence.	FRI 19-JUN-2002
ACCESSION	AC040977		
VERSION	AC040977.10		
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1 (bases 1 to 174741)	
Brizzen, B., Linton, L., Nusbaum, C. and Lander, E.	
Homo sapiens chromosome 17, clone RP11-589P10	
Unpublished	
2 (bases 1 to 174741)	
Brizzen, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,	

TITLE	JOURNAL
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (bases 1 to 174741)	REFERENCE
Birtten, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N., Campbell, A., Caser, R., Chellapandian, P., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeAngeliano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Gagen, J., Gardyna, S., Ginde, S., Goylette, M., Graham, L., Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karakas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., MCPhetters, R., Meldrum, J., Menaul, L., Mihova, T., Miranda, C., Menga, V., Morrow, J., Murphy, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Planci, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, R., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Tallamas, J., Testeigne, S., Theodore, J., Tittrell, A., Travers, M., Trisilano, J., Vassalli, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	
Direct Submission	

Anderson, S., Barua, N., Bastien, V., Bloom, T., Bogunavskiy, L., Boukfiglalier, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazarro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodds, S., Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, M., Gage, D., Galgan, J., Gardyna, S., Ginde, S., Gold, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karakas, A., Kells, C., Laocque, K., Lamazates, R., Landers, T., Lehoczy, K., Levine, R., Lindblad-Toh, K.,

TITLE	JOURNAL	REFERENCE	AUTHORS
Submitted (15-JUN-2002)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	4 (base 1 to 174741)	Bitnen,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Direct Submission			
Submitted (15-JUN-2002)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	4 (base 1 to 174741)	Bitnen,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

```

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (19-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 15, 2002 this sequence version replaced gi:21307465.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9393
Center clone name: 589_p_10
-----

FEATURES
source
    location/Qualifiers
    1..174741
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="17"
    /map="17"
    /clone_lib="RP11-589P10"
    /clone_1kb="RP11-11 Human Male BAC"
    complement(33..282)
    /rpt_family="L1MA2"
    complement(283..798)
    /rpt_family="MER9"
    complement(799..2001)
    /rpt_family="HBRVK9"
    complement(1998..6157)
    /rpt_family="HBRVK9"
    complement(6158..6667)
    /rpt_family="MER9"

```

```

repeat_region complement(6668. .8077)
/rpt_family="L1MA2"
repeat_region 8078. .8130
/rpt_family="TG)n"
repeat_region complement(8131. .8311)
/rpt_family="L1MA2"
repeat_region 8438. .8466
/rpt_family="AT_rich"
repeat_region complement(8514. .11630)
/rpt_family="L1PA4"
repeat_region complement(11632. .14356)
/rpt_family="L1MA2"
repeat_region 14357. .14587
/rpt_family="AluSc"
repeat_region complement(14588. .15568)
/rpt_family="L1MA2"
repeat_region 15658. .15846
/rpt_family="LTR33"
repeat_region complement(16982. .17028)
/rpt_family="MLT1K"
repeat_region 17739. .18034
/rpt_family="AluSx"
repeat_region complement(20741. .20968)
/rpt_family="MER20"
repeat_region 22009. .22049
/rpt_family="CA)n"
repeat_region complement(22323. .23073)
/rpt_family="L2"
repeat_region complement(25959. .26007)
/rpt_family="MIR"
repeat_region 26008. .26301
/rpt_family="AluSg"
repeat_region complement(26302. .26377)
/rpt_family="MIR"
repeat_region 27051. .27328
/rpt_family="AluY"
repeat_region 27453. .27752
/rpt_family="AluSx"
repeat_region complement(27872. .28185)
/rpt_family="AluSp"
repeat_region complement(27944. .27951)
/rpt_family="AluSx"
repeat_region complement(27963. .27981)
/rpt_family="L2"
repeat_region complement(27987. .27993)
/rpt_family="L1LJ2"
repeat_region complement(28010. .28014)
/rpt_family="L1M4"
repeat_region 28574. .28668
/rpt_family="MER91C"
repeat_region 29584. .29882
/rpt_family="AluSc"
repeat_region 30411. .30507
/rpt_family="L2"
repeat_region complement(30687. .30948)
/rpt_family="MLT1J2"
repeat_region 31119. .31420
/rpt_family="AluSc"
repeat_region complement(31462. .31759)
/rpt_family="AluSc"
repeat_region complement(31797. .31970)
/rpt_family="L1M4"
repeat_region 31995. .32019
/rpt_family="AT_rich"
repeat_region complement(32296. .32340)
/rpt_family="MLT2B3"
repeat_region 32346. .32409
/rpt_family="CATATA)n"
repeat_region complement(32411. .32699)
/rpt_family="MLT2B3"
repeat_region complement(32792. .33094)
/rpt_family="AluSx"
repeat_region complement(33104. .33179)

```

```

repeat_region complement(33249. .33789)
/rpt_family="L1MD2"
repeat_region complement(33791. .35118)
/rpt_family="L1PA13"

Query Match 9.6%; Score 67.8; DB 9; Length 174741;
Best Local Similarity 46.7%; Pred. No. 0.054;
Matches 250; Conservative 0; Mismatches 282; Indels 3; Gaps 1;

OY 31 ATATGATTTAAATGGCAATATGAGATGAGATTTTAAATTTTATATATATAA 90
129393 AAATATTTTCCTTGTGATTTAAATTTAAATTTAAATTTAAATTTAAATATA 129334
OY 91 ATATTTTTCATGAAATTTCTTTTAAAGATTAATTTGACATACAGATATATTTT 150
129333 TTTAATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTA 129274
OY 151 CC---GCACGAATTGAGTACGACAGATGTCCTGCTTTTAACTCGTAAGTTTGG 207
129273 ATATTTGAACATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 129214
OY 208 TCTTACCCAAACCAATACCAAGTAAAGCATACCGTTAATATTTCTAAC 267
129213 TAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 129154
OY 268 GGTATTAAGTTTACATTAATCATTTACTATCCGCCGTAAATTAAGTTTATGAATGT 327
129153 AATATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTA 129094
OY 328 GGTATTTTGTAGGTCAGTGAATTTTAAATTTTGTAGTCTGTTTCTTTTGG 387
129093 AATTAAGTTAAATTTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 129034
OY 388 GTTCAACTAGTACTTTTCTTTCCTTGCATCAAAATTTGTAGACGAGTGCATATA 447
129033 TTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAT 128974
OY 448 TAGATGCGAAATGAATATTTAGATTAATTTAAATTTAGAAAGCAAAATAA 507
128973 TAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTA 128914
OY 508 AATTAGTGAAAGGTACTGGAAGAAAGGCTTATGCTTGCGAAGATTCAT 562
128913 AAATTTATTAATTTAAATTTAAATTTAAATTTCTTGTAGCTTTAAAGGCA 128859

RESULT 11
CR847887 225328 bp DNA linear HTG 08-OCT-2004
LOCUS Dantio rerio clone DKEX-178N6, *** SEQUENCING IN PROGRESS ***, 7
DEFINITION unordered pieces.
ACCESSION CR847887
VERSION CR847887.2 GI:54019808
KEYWORDS HTG; HTGS PHASE1.
SOURCE Dantio rerio (zebrafish)
ORGANISM Dantio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Dantio.
1 (bases 1 to 225328)
REFERENCE
AUTHORS Sime,S.
TITLE Direct Submision
JOURNAL Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT
On Oct 9 2004 this sequence version replaced gi:53850225.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information

```



FEATURES	Location/Qualifiers
source	1. .225328

ORIG:

Ma

99

Db 111814 TATATATATATAATATATTAAATATTAATAATAAAATTTT 111865

**COMMENT**

Invert size: 191089; 11.8% error; agarose-fp

```
misc_feature      1. 16123
                  /note="assembly_fragment:01275
                  fragment_chain:1"
misc_feature      16224. 26995
                  /note="assembly_fragment:00826
                  fragment_chain:1"
misc_feature      27096. 29762
                  /note="assembly_fragment:00044
                  fragment_chain:1"
misc_feature      29663. 32823
                  /note="assembly_fragment:00143
                  fragment_chain:1"
misc_feature      32924. 59473
                  /note="assembly_fragment:01558
                  fragment_chain:2"
misc_feature      59574. 77455
                  /note="assembly_fragment:01037
                  fragment_chain:2"
misc_feature      77556. 80898
                  /note="assembly_fragment:00073
                  fragment_chain:2"
misc_feature      81090. 84886
                  /note="assembly_fragment:00229
                  fragment_chain:3"
misc_feature      84587. 87803
                  /note="assembly_fragment:00165
```

```

VERSION AC11559.2 G1:28828893

```

KEYWORDS HTG.  
SOURCE Dictyostelium discoideum  
ORGANISM Dictyostelium discoideum  
REFERENCE 1 (bases 1 to 125623)  
AUTHORS Gloeckner, G., Eichinger, L., Szefranek, K., Pachbat, J., Dear, P.,  
Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kump, K.,  
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A., and  
Noegel, A. A.  
Sequence and analysis of chromosome 2 of Dictyostelium discoideum  
Nature 418 (6893), 79-85 (2002)  
22092622  
MEDLINE  
JOURNAL 12097910  
PUBMED  
REMARK The Dictyostelium Genome Sequencing Consortium  
REFERENCE 2 (bases 1 to 125623)  
AUTHORS Baumgart, C.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
3 (bases 1 to 125623)  
AUTHORS Baumgart, C.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
4 (bases 1 to 125623)  
AUTHORS Baumgart, C.  
TITLE Direct Submission  
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
On Mar 4, 2003 this sequence version replaced gi:19570032.  
CDS predictions from GeneID do not necessarily reflect true genes.  
Further information is available from IMB Jena, Department of  
Genome Analysis  
(http://genome.imb-jena.de/dictyostelium/)  
and the University Cologne, Institute for Biochemistry I  
(http://www.uni-koeln.de/dictyostelium/project.shtml  
Funding  
Agency : Deutsche Forschungsgemeinschaft (DFG).  
Location/Qualifiers  
1. 125623  
/organism="Dictyostelium discoideum"  
/mol\_type="genomic DNA"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/chromosome="2"  
/map="4229038-4354721"  
join<416..688,837..1349,1434..1493,1638..2931)  
/note="Geneid exon scores (in order of location ranges) :  
10.12, 21.67, 1.80, 118.27 - GSCJ\_ID dd\_00183"  
/codon\_start=2  
/product="similar to Dictyostelium discoideum (Stime  
mold). GP138"  
/protein\_id="AA192328.2"  
/db\_xref="GI:2882894"  
/translation="FKGNQKQOQVALLSQIYQVYEGFQCLMNMVSGDVTIGINI  
FGVPTLTVDSFEDLTFISSLAISPSNIFSLPNTSPSEVFENIGKNDQIE  
DIVLENTLVTFKSIQVLPNAPFSTGALITLLINANPAPLPQOPNPNTKIDN  
LNIINGPSPNAGSALSNLTETLNTNNYNTKNTFFPPEPNOITKALINPHSDP  
LNQLEFPEPNNINPLELHLENLGNISNDVLTLSKNLTLTSHNLLPNKPY  
PLNCSLDLSNPVYSIVDVRNLTSLSNVYGENLPASNPDEPKLYLLLSNQF  
NNSLPEEYCSFVKSIALGSDNLTGSVPCFKLGEPEPGLFPNNFDPFDPADV  
CAFSLDVFNQLKTNETTPEFTSGNSLGMDSIDSTQANVELAIENKKIKITIP  
PAGIONATLKFGNSPSEIIVINERYEPTIDSYVQSGIFPFGSFGSYNDYENIF  
TINKYKTAHQTIEDQNLNNGVGENDPISVLKRNFAVNSIDVAGKQSEYTF  
ITFGNTLSDLSIQNLNTGDFSGFGCTTDFSNPIKINDIDVQILLETETR  
INITYPLIENAGTYLIEVGDPSFNEIKFETPPPTPTPTSTPTSTSTST  
SDDEGISTPTLSISRIYLSIILLINPYF"  
complement(join(3950..4537,4680..4808))  
/note="Geneid exon scores (in order of location ranges) :  
38.84, 6.50 - GSCJ\_ID dd\_00185"  
/codon\_start=1  
/product="similar to Ralstonia solanacearum (Pseudomonas  
solanacearum). Putative GST-related protein (EC 2.5.1.18)"

/protein\_id="AA051485.1"  
/db\_xref="GI:28828894"  
/translation="MQNKFDEFTIINDHKSPTKYVELMLIELGVSYESISINDYLN  
NSNRIPALIDMTFPKVPNTLYSGSVLILAKFSSVLLDPLNATLQNNQSIISKS  
VMQVYLIPSPNN  
ADISVWLLHLLKNSNQIISRSNINKMELINORDSIKLLNNNNNNNNNNNN  
NNNNNGIKFSNDGLAINAL"  
complement(5354..7216)  
/note="Geneid exon scores (in order of location ranges) :  
110.35 - GSCJ\_ID dd\_00186"  
/codon\_start=1  
/product="similar to Plasmodium falciparum (isolate 3D7).  
Hypothetical protein"  
/protein\_id="AA051486.1"  
/db\_xref="GI:28828895"  
/translation="MGFTMEYENNRKQDGVDDNNNNNNNNNNNNNNNNNNNN  
YSSSSNNNNNNNDYSDNDYNOITPIYQNDYQNDSTNNNNNNNNNDYNOYPLN  
NDFQDTNNNSNTYDYNNYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NOINNSNN  
IYKNSNLGNKNDINEKSNKYEPYKQDSYTIKQDIGNI VILFKQACAYQSDLI  
IAIKQSKTDILKEIKFVGLDYKESISTLLFKMIINDKTAKEPFEYNPSPESTI  
LIPESDPELLKPNHLLIQSGSPPIHRRKMPPTKFTSTYPTGTTEPKYKQGPV  
KIDRKCEYVIDIKFSQDSIPIQYGFYKMGVSEQITFNYSPIREPKPIYHNDN  
NNSERNATWLNFLILFERYLQSRVNDGRTVYGOPLNVLVSQSSIEILPLKQSI  
NOSKLCIGRSTLSFGNPLCNTHYSKDLMDAEKCPYFLDANNEPIKQLPDQLES  
FKNEKKEKKEVFIKDFAYIDPSNLOSILDKKVPYQENNNYNNNSYDYDY"  
complement(7921..9687)  
/note="Geneid exon scores (in order of location ranges) :  
222.87 - GSCJ\_ID dd\_00188"  
/codon\_start=1  
/product="similar to Agrobacterium tumefaciens (strain C58  
/ATCC 33970). 3-methylcrotonoyl-CoA carboxylase beta  
subunit"  
/protein\_id="AA192331.2"  
/db\_xref="GI:28828896"  
/translation="MLKSIISLKNQIILKNIINNGRILNNVGEKLSKSKLLKINYS  
STTDRTFNLDGTDIDKNSAEYKDNLIINNSTKQLEKENTIKLGGEGLANDKINSRG  
KLIVBERIALIDVGPFLFESQLAGMGWYGEVAAAGIITIGIGIHEVACVYAND  
STVKGGTPIPIVKKHILROEITQENNNLPCITLVGSGANLRODVPDRHFGPIF  
PNOAMSAKRIPQIIVWMSCTAGAYYPAMADESVIYKGTITPILGSGPIYKAAIGE  
IYTSBELGADJHCTSGVTDHYARDDBAIAITRIYVSNLRKRPQSPVITETEPPL  
YPTSLAGIYVPSDLKKNPDIRYIARLVAGVDFEKLXGTLLCGFARVGMGMV  
IANNGILFESAVKGAHFIELCNQIGIPLVFLQNTIGFWGTVESGIAKQCAVAMV  
AAVATKVPKITWIIIGSGFAGNYGMCGRSFRPLPYMWNNAISVNGGQAASVLAQI  
OKDMAKKNKQWSPREENTFKKPIDSKPEEBSIYYSARCMDDVIDPQDSRKVAL  
SISACMNPINPSPSGFVFRK"  
complement(9983..11383)  
/note="Geneid exon scores (in order of location ranges) :  
98.71 - GSCJ\_ID dd\_00191"  
/codon\_start=1  
/product="similar to conserved hypothetical protein  
[Schizosaccharomyces pombe]"  
/protein\_id="AA192333.2"  
/db\_xref="GI:28828897"  
/translation="KRATHSREVDVTESTINIKETSTTKVTTESAIDRFKSKSKTK  
KHKHVKVSSDNDNDHNSINNERKVKKSTLSLKNKKPSIKKDLBPVNDORDT  
AKRAEIKYLENDRLBEGNPEYKMLETQYKRYKLANGELRLGSLVDINTLSERV  
CYRDTITLKLEVYKETSINKGEIKETFGSSQFLAKSDVDITLWDSFETLT  
KNDITTKCYQSSILIRANGFYNIKPIIHAKPIYKFPDKPEFCDITLTSQNTG  
VYKECOLPILPVLIIIFCKMASVLIINDASQGLSSYSITNNVI FYLQKGLPLSY  
KOIQGFKSSKEVQVYVNNRDADLLVYEFKGYIFDVRNMMVILVNDNKKEDPD  
AEEELKPNRRYDKNHNYLPVDMKDLDTDLFLIKDCIAHNTTRCILRSQPLTLI  
KYFOKAYNLISGKLP"  
13531..11383  
/note="Geneid exon scores (in order of location ranges) :  
16.95 - GSCJ\_ID dd\_00194"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AA192334.2"  
/db\_xref="GI:28828898"  
/translation="WSIITYSPVPFQLEFVELRDLDVNSEERKFAVQIFNVGVPSTIK  
KIQVYVNSVWKNYKNLGNLSIKELIFVHGRSLNPKLIYTNGLIVERTSAGLYPA"  
14573..14956



*	47133	47233:	gap of 100 bp
*	47233	:	cont'g of 31791 bp in length
*	79024	79123:	gap of 100 bp
*	79124	:	cont'g of 52360 bp in length
*	131484	131583:	gap of 100 bp
*	131584	:	cont'g of 16811 bp in length

**ORIGIN**

```
Query Match      9.5%; Score 66.8; DB 2; Length 148394;
```

QY	19	AAATTTTAAATATATGTAATTTAAAGCCAAATAAAGTTTGAAGAGAGTTT	78
Db	78625	ATATTTGTTATATAATTAATAATTAATTAATTTTAAATTTATATTTTAAAT	78566
QY	79	TTTATATATAAATTAATTTTTCATGCAATTCGTTTAAAGATAAAATTTTGACATTAAC	138
Db	78565	TTATATTTATTTAAATATATAGTATTAATTTTAAATTTAATTAATATAATATATAAT	78506
QY	139	AAGATATATTTCCGACAGATTTGAGTCACAGAGATGTCCTGTCCTTAAACCTCGTA	198
Db	78505	AATTTATTTATTTATTTTAAATATATTAATAATTAATAATTAATTAATTAATAATA	78450
QY	199	AAGTTTGGTCTAACCCAAACCAATACCAAGGAAGAAAGACATACCGGTATTAAT	258
Db	78449	TGATATTTATTTATTTATATTTTATTTTAAATTTTATTAATTTTAAATTAAT	78390
QY	259	ATTCTAACCGGTTATTAAGTTTACATTAATCATTTTACTAATCCGGGTATTAAGTTT	318
Db	78389	ATTTAAATAATTTATTAATAATATATTAATTTTAAATTTAAATTTGAATTAATTAATTT	78330
QY	319	ATGAATATGCTAATTTTGTAGTGCAAGTAATTTATTAATTTTATAGTATCTGTTTT	378
Db	78329	TATATATTTTATATTTTTTTATTAATTAATTTTATTTTAAATTAATTAATTTTATAT	78270
QY	379	CTTTTGGGTTCAACTAGTTTACTTTTTCCTTTGCAATCAAAATTTGTAGACGAGTG	438
Db	78269	ATATATTTATTAATTTATTTAAATAATGTAATTTTATATTTATTTTATTTTAAATA	78210
QY	439	GTCATATATATAGTGTGAATGAATAATTTGATATAATAATAATAGA-AAGTG	497

Dd	78209	TT - ATTTTAAATTAATTAAATTTAATAATTTATTTATTAATATAATATAATTAATTT	78152
Oy	498	AACAAAAAATTAAGTGAAAAGCTAACGGAAGAAAAGCGTTATGGCTTGCAAGAT	557
Dd	78151	TATATATTTTAATTAATATATATATATATATATTTAATATATATAATTTATAT	78095
Oy	558	TCATATAGTTTATTTGCACCAAAAAGAGAAGACTGGCCTGTCTCTCTCTTTTA	617
Dd	78091	AAATTTATTTTTTTTATATAAAAATATATTTTAATTATATATATTTTTTAATTATA	78032
Oy	618	AAATGCTPAA 627	
Dd	78031	AAAAATPAA 78022	

RESULT	15
CR759796/c	
LOCUS	
DEFINITION	Cr759796 Danio rerio clone DKEx-18A10, WORKING DRAFT SEQUENCE, 10 uncoreted pieces.
ACCESSION	CR759796
VERSION	CR759796.4 GI:52673634
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Danio rerio (zebrafish)
ORGANISM	Danio rerio Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE	1 (bases 1 to 196762) Burton,J
AUTHORS	Direct Submission
TITLE	Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL	zfish-helpp@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 24, 2004 this sequence version replaced gi:52626701.
COMMENT	

\* 109884 109883: gap of 100 bp  
\* 109984 118156: contig of 8173 bp in length  
\* 118157 118256: gap of 100 bp  
\* 118257 129072: contig of 10816 bp in length  
\* 129073 129172: gap of 100 bp  
\* 129173 182677: contig of 53505 bp in length  
\* 182678 182777: gap of 100 bp  
\* 182778 196762: contig of 13985 bp in length.  
\* Location/Qualifiers

## FEATURES

Source  
1. 196762  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEX-18A10"  
/clone\_lib="Daniokey"  
1. 5760  
/note="assembly\_fragment:00009  
fragment\_chain:1"  
5861. 20214  
/note="assembly\_fragment:00500  
fragment\_chain:1"  
20315. 34680  
/note="assembly\_fragment:00347  
fragment\_chain:1"  
34781. 43318  
/note="assembly\_fragment:00123  
fragment\_chain:1"  
43419. 73325  
/note="assembly\_fragment:00837  
fragment\_chain:1"  
73426. 109883  
/note="assembly\_fragment:01226  
fragment\_chain:1"  
109984. 118156  
/note="assembly\_fragment:00051  
fragment\_chain:1"  
118257. 129072  
/note="assembly\_fragment:00216  
fragment\_chain:2"  
129173. 182677  
/note="assembly\_fragment:01642  
fragment\_chain:2"  
182778. 196762  
/note="assembly\_fragment:00664.0"

## ORIGIN

Query Match 9.4%; Score 66.6; DB 2; Length 196762;  
Best Local Similarity 48.5%; Pred. No. 0.08;  
Matches 242; Conservative 0; Mismatches 254; Indels 3; Gaps 2;

QY 11 ATTTGGGAAAATTAAATATATATGATTAATGCAATAAAGTTAGATGAGATTTT 70  
Db 118101 ATTTAAATTAATTTATTAATTTATATATATATATATATATATATAT 118042  
QY 71 TTTAAATTTTATATTAATAATTTTTCATGAAATGTTTAAAGTAAATTTTG 130  
Db 118041 TTTTATTTTATTAATTAATTAATTTTAAATATATATATTTTATTTATATAT 117982  
QY 131 ACATAACCAAGTATTTTCCGCCAAGATTGAGTCTACGAGAGATGTCCTTTAA 190  
Db 117981 ATTTATATTTTATTTATATTTATTAATAATTAATTAATTAATTAATTA 117922  
QY 191 ACCTCGTAAAGTTTGCTTACCAACCAATACCAACCAAGTAAAGCAATACCGG 250  
Db 117921 TTTTAT -AATTTATATATTAATTTATTTTATTAATTAATTTTATATATTA 117864  
QY 251 TTTAATAATTTCAACCGTTTAAAGTTTACATTAATCATTTACATCCGCTGAAT 310  
Db 117863 TTTATTAATTAATTTTATTAATTAATTAATTAATTAATTTTATATATTTAT 117804  
QY 311 TAAGTTTATGAATGCTTATTTTGTAGTCAAGTAAATTTATTAATTTTATAGTAC 370  
Db 117803 TAAATATATATATATATATATATTTTAAATTAATTAATTTATTAATTAATTA 117744

QY 371 TTTGTTTTCTTTTGGGTTCACTACTTCTTTTCCTTGACATCAATTAATTTGTA 430  
Db 117743 TAAATATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTTAT 117684  
QY 431 GACAGTGGTCCATATATATGATGCGTGAATGAATGAATTTGATTAATTAATTA 490  
Db 117683 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 117625  
QY 491 GAAAGTGAACAAAAAAAT 509  
Db 117624 TATATATATATATATATAT 117606

Search completed: November 17, 2005, 16:54:19  
Job time : 3275 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2005, 13:43:52 ; Search time 473 Seconds

(without alignments)  
8823.293 Million cell updates/sec

Title: US-10-810-788a-4

Perfect score: 705  
Sequence: 1 aagcttcaattcgggaaa.....tctgtggtcacttgatacc 705

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: geneseqn1980s:\*\n2: geneseqn1990s:\*\n3: geneseqn2000s:\*\n4: geneseqn2001as:\*\n5: geneseqn2001bs:\*\n6: geneseqn2002as:\*\n7: geneseqn2002bs:\*\n8: geneseqn2003as:\*\n9: geneseqn2003bs:\*\n10: geneseqn2003cs:\*\n11: geneseqn2003ds:\*\n12: geneseqn2004as:\*\n13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.2	9.1	8056	8	ABZ10246 Haematopo
2	63.4	9.0	612	4	AAH71471 Human cer
3	62.6	8.9	8056	8	ABZ10246 Haematopo
4	62.4	8.9	531	13	ACNS1594 Cotton an
5	62.2	8.8	8056	8	ABZ10100 Haematopo
6	60.6	8.6	11422	6	ABK31936 Human che
7	60.6	8.6	11422	6	ABL32218 Human imm
8	60.6	8.6	560	13	ACN47472 Cotton pr
9	60.2	8.5	626	5	ABV60941 Human pro
10	60	8.5	499	13	ACN62049 Cotton gy
11	59.4	8.4	8056	8	ABZ10100 Haematopo
12	59.4	8.4	15954	6	ABK31495 Signal tr
13	59.4	8.4	15954	6	ABL70468 Chemical1
14	59.2	8.4	6816	12	ADQ24856 Human soc
15	59.2	8.4	50000	6	ABL55644 Human soc
16	59	8.4	469	13	ACN62049 Cotton gy
17	59	8.4	752	5	ADL44766 Human ova
18	58.8	8.3	428	13	ACNS2919 Cotton an
19	58.6	8.3	469	13	ACN62049 Cotton gy
20	58.6	8.3	547	13	ACN62109 Cotton gy

C	21	58.4	8.3	631	5	AD173251 Human ova
C	22	58.4	8.3	631	5	AD173251 Human ova
C	23	58.2	8.3	554	13	ACNS1258 Cotton an
C	24	58	8.2	9646	6	ABL33688 Human imm
C	25	57.8	8.2	494	5	ABV10021 Human pro
C	26	57.8	8.2	7823	4	AA545490 Chemical1
C	27	57.8	8.2	7823	4	ABL34061 Human imm
C	28	57.8	8.2	7823	6	ABK31493 Signal tr
C	29	57.8	8.2	7823	6	ABK28418 DNA trans
C	30	57.6	8.2	353	13	ACNS3258 Cotton an
C	31	57.6	8.2	466	13	ACN61227 Cotton gy
C	32	57.4	8.1	550	5	ABV57072 Human pro
C	33	57.4	8.1	565	13	ACNS6344 Cotton an
C	34	57.4	8.1	4857	13	AD589407 Oligonuc1
C	35	57.4	8.1	7857	6	ABQ67075 Human ang
C	36	57.4	8.1	9001	13	AD589773 Oligonuc1
C	37	57.2	8.1	353	13	ACNS3258 Cotton an
C	38	56.8	8.1	8423	6	ABL33407 Human imm
C	39	56.6	8.0	376	5	ADL44697 Human ova
C	40	56.6	8.0	574	13	ACNS6366 Cotton an
C	41	56.4	8.0	385	5	ADL37765 Human ova
C	42	56.4	8.0	385	5	ADL72626 Human ova
C	43	56.2	8.0	17848	4	AA545323 Chemical1
C	44	56.2	8.0	17848	4	ABK39976 Human che
C	45	56.2	8.0	17848	6	ABK28164 DNA trans

## ALIGNMENTS

RESULT 1  
ABZ10246/c  
ID ABZ10246 standard; DNA; 8056 BP.

AC ABZ10246;  
XX 16-JAN-2003 (first entry)

XX 16-JAN-2003 (first entry)

XX Haematopoietic cell proliferation disorder related DNA sequence #386.

XX Human, haematopoietic cell proliferation disorder; cytostatic;

XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;

XX cytosine methylation state; gene; ds.

XX Homo sapiens.

XX W0200277272-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WC-EP003401.

XX 26-MAR-2001; 2001US-0278333P.

XX (EPIC-) EPICENOMICS AG.

XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
PI Olk A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Ieu F;  
PI Lewin A, Lipscher E, Mader S, Model F, Mueller V, Otto T, Pelet C;  
PI Schwope I, Ziebarth H;

XX WPI; 2003-018942/01.

XX Detecting and differentiating between hematopoietic cell proliferative

XX disorders, comprises contacting a target nucleic acid with a reagent that

XX distinguishes between methylated and non-methylated CpG dinucleotides.

XX Claim 28; SEQ ID NO 386; 117pp; English.

XX The present invention describes a method for detecting and

XX differentiating between hematopoietic cell proliferative disorders

XX associated with at least 1 gene and/or their regulatory regions in a

XX subject. The method comprises contacting a target nucleic acid in a

biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118 CC represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used: for differentiating between healthy haematopoietic cells and proliferative disorder haematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables a highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients

Query Match	9.1%	Score	64.2	DB	8	length	8056
Best Local Similarity	45.7%	Pred.	No. 0.022				
Matches	260	Conservative	0	Mismatches	308	Indels	1
						Gaps	1

QY	5	TTTTAAATTTGGGAAATTTTATATAATATGTAATTAATGGCAATTAAGGTTACATGAG	64
Db	7154	TTACAAATTATATATATATATATTTTATTAATATTTAAAAAAATTAATTAATTAATAT	7095
QY	65	AGTTTTTTTAAATTTTATATATAAAATTAATTTTTTGCATGAATGTGTTTTAAATATAA	124
Db	7094	TAAAAAATTATATTTAAAAAATATTAATTTTAAATTAATTAATTAATTTTTTTTTTTT	7035
QY	125	ATTTTGACATACCAAGTATTTATTTCCGCCACGAATTGAGCTACGAGAGTGTCTGT	184
Db	7034	AAAAAAAAATCAAAAATTAATTTTTTTTTTTTTTTTTTAAATTAATAATCAATATATTTTT	6975
QY	185	CTTAAACCTGCPAAGTTTGGTCTTCCCAACCCAAATCCCAAGTAAGTAACGACCA	244
Db	6974	TATATATATATCATTTTTTAAATTAATAAAAAATAAAAAAATTTAATAAAAAATTAACA	6915
QY	245	TACCGTTAATATATATCTAAACCGTTATATAGTTTACATTAATCATTTACTTAATCCCG	304
Db	6914	TTCAATTTATTCAAAAAATAAATTAACAATTTTTTTTTTAAATAATTTTTTATTAATTT	6855
QY	305	TGTAATTTAGTTTATGCAATGTGCTATTTTGTAGTCACTGCAATTTATTAATTTTT	364
Db	6854	TTTCAATTCATTAATAATTAATTTTTTATTTATTTTAAAAAAATTCATTTTCACAATTC	6795
QY	365	TAGTACTGTCTTCTTTTTGGGTCAACAGTACTTTTTTCTTGACATCAAAAT	424
Db	6794	AAATATTTTTTATTTTTTTTATTTTTTTCATTTAATATATTTTTTAAATTTTAAATCA	6735
QY	425	ATTGTAGACGAGTGTCATATATAGATGCGAATGAATGAATATTTAGATTAATAATA	484
Db	6734	ATATTTCAAATATATATTAATCAACATTAATTAACATTAATAAAAAAATTTATTTTTTAA	6675
QY	485	AAATATGAAAGTGAACAAAAAAATTAAG-TGGAAAAGGTAACTGGAAAAGAAAAGCTTAT	543
Db	6674	AAATTAATTAATAAAATTTAAAAATTCATATTAATAAAACATTAATAAACAATAAATAT	6615
QY	544	TGGCTGGCAAGATTCATTAAGTTATTT 572	
Db	6614	AAATTAATTAATAAAATTAATAATTTTTTTT 6586	

RESULT 2  
AAH71471/c  
ID AAH71471 standard; cDNA; 612 BP.

AC	AAH71471;
XX	
DT	19-SEP-2001 (first entry)

XX		
DE	Human cervical cancer marker nucleic acid 2745.	
XX		
XX	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200142467-A2.	
XX		
PD	14-JUN-2001.	
XX		
PF	08-DEC-2000; 2000WO-US033312.	
XX		
PR	08-DEC-1999; 99US-0169681P.	
PR	21-DEC-1999; 99US-0171350P.	
PR	14-MAR-2000; 2000US-0189315P.	
PR	12-MAY-2000; 2000US-0203791P.	
PR	09-JUN-2000; 2000US-0210600P.	
PR	21-JUL-2000; 2000US-0220114P.	

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Deeds J, Berger A, Zhao X;

DR WPI; 2001-375006/39.

**PT** New isolated nucleic acid for diagnosing and treating cervical cancer and  
**PT** for assessing and detecting compounds for treating the cancer.

PS Claim 1; Page 564; 1051pp; English.

CC The invention relates to novel genes (AAH68777-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy

SQ Sequence 612 BP; 380 A; 50 C; 2 G; 178 T; 0 U; 2 Other;

Query Match	9.0%	Score	63.4	DB	4	Length	612
Best Local Similarity	52.7%	Pred. No.	0.029				
Matches	136	Conservative	0	Mismatches	122	Indels	0
						Gaps	0

Qy	251	TTAATAATATCTCAACCGTTATAAGTTACATAAATCATCTTACATAACCGCGTAAAT	311
Db	265	TTTTTTTTTTTAAAAATATTTTTTGGTATGTTAAATTAATGGATGGGTGATTTT	206
Qy	311	TAAGTTTATGAATCTGGTTATTTTGTAGGTCAAGTAAATTTATATTTTGTAGTAC	376
Db	205	TATATTTTTTTTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAT	146
Qy	371	TTGGTTTTTCTTTTTGGGTCAACAGTACTTTTTTCCTTGACATCAAAATATATTGA	433
Db	145	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTAATAAAAAATTTAAA	86
Qy	431	GACAGATGGTCATATATAGATGTGAATGAATGAATATTTGAGTAAATTAATTAATA	490
Db	85	AAA	26
Qy	491	GAAAGTGAACAAAAA	508
Db	25	AAAAAAAAAAAAAAAAAAAA	8

RESULT 3  
ABZ10246  
ID ABZ10246 standard; DNA; 8056 BP.

AC ABZ10246;  
XX





PS Claim 1; SEQ ID NO 6375; 34bp; English.  
 XX  
 XX The invention relates to 17880 cotton expressed sequence tags (ESTs;  
 CC ACN45220-ACN61099). The ESTs were isolated from cDNA libraries generated  
 CC from primed or non-primed seeds from variety DP50B, mature seeds from  
 CC variety Coker 312 Boswell 96 field, and androecium tissue, gynoecium  
 CC tissue, developing fibres, carpel walls and septa from variety  
 CC Nucleon33B. The invention also relates to substantially purified  
 CC proteins or their fragments encoded by nucleic acid molecules of the  
 CC invention, and to transformed plants having a nucleic acid construct  
 CC comprising a nucleic acid of the invention. The cotton ESTs are useful as  
 CC molecular tags to isolate genetic regions, to isolate genes, to map  
 CC genes, to determine gene function and to determine whether genes are  
 CC members of a particular gene family. The nucleic acid molecules may be  
 CC used for isolating a variety of agronomically significant genes  
 CC associated with plant growth, quality, yield, and could also serve as  
 CC links in metabolic and catabolic pathways. The nucleic acid molecules are  
 CC also useful for identifying genes important in initiating and maintaining  
 CC seed germination or that may be used to mitigate stresses encountered  
 CC during seed germination. The ESTs additionally enable the acquisition of  
 CC promoters and cis-regulatory elements which will be useful to express  
 CC agronomically significant genes in these tissues and/or other tissues,  
 CC and also permits the acquisition of molecular markers useful in breeding  
 CC schemes, genetic and molecular mapping, and in cloning of agronomically  
 CC significant genes. The nucleic acid molecules are further useful for  
 CC detecting the expression level or pattern of a protein or mRNA and for  
 CC detecting the presence or quantity of a protein by tissue printing. The  
 CC present sequence represents a specifically claimed EST isolated from a  
 CC cotton variety Nucleon33B androecium tissue cDNA library (LIB3828). The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was patented in electronic format directly from the US  
 CC Patent Office at [seqdata.uspto.gov/sequence.html?DocID=US20040123340](http://seqdata.uspto.gov/sequence.html?DocID=US20040123340)  
 CC  
 XX  
 SQ Sequence 531 BP; 287 A; 0 C; 70 G; 172 T; 0 U; 2 Other;

Query Match 8.3%; Score 62.4; DB 13; Length 531;  
 Best Local Similarity 51.4%; Pred. No. 0.043;  
 Matches 144; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 315 TTTTAAAGAAATGTTATTTGTAGTACGGAATTTATTTTATTTTACTGTT 374  
 Db 1 TTTTGTGTT 60  
 QY 375 TTTTCTTTTGGGTTCACTAGTACTTTTTCCTTGACATCAAAATTTGTAGACG 424  
 Db 61 TTTTCTTTTGGGTTCACTAGTACTTTTTCCTTGACATCAAAATTTGTAGACG 120  
 QY 435 AGTGGTCCATATATAGTGTGAATGAATGATATGATTAATAATATATGAAA 494  
 Db 121 AAAAAAAAAAAAAAAAAATATATAAGTAAATGGGTTTGAAAAAAAAAAAAAGAAA 180  
 QY 495 GTGACAAAAAAATATAGTGAAGAGTAACTGAAAAGAAAGCTTATTTGGCTGGCA 554  
 Db 181 AAAAAAAAAAAAAAAAAATATATAAGTAAATGGGTTTGAAAAAAAAAAAAAGATTTGGGG 240  
 QY 555 GATTCCATAGTTTATTTTCAACAAAAAGAAAGATGACTT 594  
 Db 241 GGATTAAAAAGGTTTTTAAAAAATTTTAAAAAATTTTAAAAAATT 280

RESULT 5  
 AB210100/c  
 ID AB210100 standard; DNA; 8056 BP.  
 XX  
 XX AB210100;  
 XX  
 XX 16-JAN-2003 (first entry)  
 XX  
 XX Haematopoietic cell proliferation disorder related DNA sequence #240.  
 XX  
 XX Human; haematopoietic cell proliferation disorder; cytostatic;  
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KW cytosine methylation state; gene; ds.

XX  
 OS Homo sapiens.  
 XX  
 PN MO200277272-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002MO-EP003401.  
 XX  
 PR 26-MAR-2001; 2001US-0278333P.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Berlin K, Braun A, Dietler J, Gietig D, Howe A, Mueller J;  
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;  
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;  
 PI Schwabe I, Ziebarth H;  
 XX  
 DR WPI, 2003-018942/01.  
 XX

PT Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent that  
 PT distinguishes between methylated and non-methylated CpG dinucleotides.  
 PT  
 PS Claim 28; SEQ ID NO 240; 117bp; English.

XX The present invention describes a method for detecting and  
 XX differentiating between haematopoietic cell proliferative disorders  
 XX associated with at least 1 gene and/or their regulatory regions in a  
 XX subject. The method comprises contacting a target nucleic acid in a  
 XX biological sample obtained from the subject with at least 1 reagent,  
 XX which distinguishes between methylated and non-methylated CpG  
 XX dinucleotides within the target nucleic acid. AB209861 to AB211118  
 XX represent specifically claimed nucleotide sequences from the present  
 XX invention. Oligonucleotides from the present invention can be used: for  
 XX differentiating between healthy haematopoietic cells and proliferative  
 XX disorder haematopoietic cells; for differentiating between acute  
 XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 XX determining the cytosine methylation state and/or single nucleotide  
 XX polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 XX related sequences and their complements; and as primers for the  
 XX amplification of haematopoietic cell proliferation disorder related DNA  
 XX sequences. The nucleotide sequences from the present invention can also  
 XX be used for detecting a predisposition to, differentiation between  
 XX subclasses, diagnosis, prognosis, treatment and/or monitoring of  
 XX haematopoietic cell proliferative disorders. The present method enables a  
 XX highly specific classification of haematopoietic cell proliferative  
 XX disorders allowing for improved and informed treatment of patients  
 XX  
 SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match 8.8%; Score 62; DB 8; Length 8056;  
 Best Local Similarity 46.1%; Pred. No. 0.054;  
 Matches 248; Conservative 0; Mismatches 285; Indels 5; Gaps 1;

QY 5 TTTTAAATTTGGAAAAATTTAATATATATATGATTAATGCGAATAAAAATTGATGAG 64  
 Db 1624 TTTCGATTTTAAAAATTTAAAAATTTAATATATATTAATTAATTAATTAATA 1565  
 QY 65 AGTTTAAAAATTTTATATATATATATATATATTTTGCATGAATGTTTAAAGTAAA 124  
 Db 1564 ATTTTAAAAAAATTAATTTTATTTAAAAATTTTAAACGATTTATATTTTAAATTT 1505  
 QY 125 ATTTGACATTAACCAAGTATTTATTTTCCGCCAGATGAGTCTACAGAGATGTCCTGT 184  
 Db 1504 TATTAATTAATTAATTAATTAATTTATTTAATATATTTTAAATAATTTTAA 1445  
 QY 185 CTTTAAACCTCGTAAGTTTGTCTTACCAACCAATATCCACCAAGGTAACACCA 244  
 Db 1444 TTTTCTTTTAAAAAATTAATTAATTAATTAATTAATTAATTTTATTTTAA 1385  
 QY 245 TACCGGTTAATATATTTCAACCGGTTTAAAGTTTAACTTAATTCATTAATCCGCG 304

Db 1384 AAAAAAAAAATTTTATATATAAAAAATTAATTTT-----TAAATATTTTATATATAAAA 1330

Oy 305 TGTATATTAGCTTTATAGAAATGCTGTTATTTTGTATAGGTCAGCTGAAATTTATTAATTTT 364

Db 1329 TTATATATATAATTTTTTTATATAATTTATATTTTTTTAAAAAATATATACGTTTTTATATATTA 1270

Oy 365 TAGTACTGTGTTTTCTTTTTGGGTTCACTAGTACTTTTTCTTTGACATCAAAAT 424

Db 1269 ATATATATTACGTAAAGTTTTTTTTTATTCGAAAAAATATTTTATATTAATTAATTAATTA 1210

Oy 425 ATTGTAGACGAGTGGTCCATATATATAGATGTGAAATGAAATGAAATATTTAGTAAATTAATA 484

Db 1209 ATTAATAATATATTAATAATTTTAAAAATTAATTAATATATATGAAAAATATATTAATTTTAA 1150

Oy 485 AATATAGAAATGTAACAAAAAATTTAGTGGAAAAAGTACTGAAAGAAAGCCTTA 542

Db 1149 ATTTTCGAATATCGAAAAAATATTTAAAAAAGCAATATATTAACGTTTTTAAAAACGTTAA 1092

RESULT 6

ABK39936

ID ABK39936 standard; DNA; 11422 BP.

XX

XX ABK39936;

DT 21-MAY-2002 (first entry)

XX

DE Human chemically pretreated gene sequence #9 strand 1.

XX

XX Human; day; bisulphite treatment; CpG; DNA methylation; cancer; tumour;

KM cytosine; ALDH6; CYP11A; CYP11B; CYP2A3; DPYD; EPHX2; OCLN; TXNRD1;

KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

OS Homo sapiens.

XX

XX WO200202806-A2.

PD 10-JAN-2002.

XX

XX 29-JUN-2001; 2001WO-EP007470.

PF

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX

PA (EPIG-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-154757/20.

XX

XX

XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,

PT useful for detecting cytosine methylation state of genes associated with

PT pharmacogenomics and for therapy of diseases e.g. cancer.

PS Claim 1; SEQ ID NO 17; 24bp; English.

XX

XX The invention relates to a nucleic acid comprising a sequence at least 18

CC bases in length of a segment of the chemically pretreated DNA of genes

CC associated with pharmacogenomics according to one of the sequences of the

CC genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B (NM\_000497), CYP3A3

CC (NM\_000776 and NM\_017460), DPYD (NM\_000110), EPHX2 (NM\_001979), OCLN

CC (NM\_002538), TXNRD1 (NM\_003330), UGT8 (NM\_003360), MRP (NM\_004996,

CC NM\_019900, NM\_019901, NM\_019902, NM\_019862, NM\_019898, NM\_019899) and

CC their complementary sequences, or a sequence (S1) chosen from 87

CC sequences and their complements. The chemical pretreatment is bisulphite

CC treatment to convert cytosines (but not methyl-cytosines) into uracils.

CC Also included are an oligomer (II) in particular an oligonucleotide or a

CC peptide nucleic acid (PNA)-oligomer, comprising an oligonucleotide or a

CC base sequence having a length of 9 nucleotides which hybridises to or is

CC identical to a chemically pretreated DNA of genes associated with

CC pharmacogenomics and their complements, arranged in an array for

CC analysing diseases associated with the methylation state (CpG) and/or

CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The

CC	oligonucleotides may also be used as PCR primers. The set of 87 nucleic acids
CC	and their complements is useful for diagnosis and therapy of solid
CC	tumours and cancer. The present sequence represents one the 87 DNA
CC	sequences or its complement. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pcr_sequences
XX	
XX	Sequence 11422 BP; 3867 A; 59 C; 1786 G; 5710 T; 0 U; 0 Other;
XX	
XX	Query Match 8.6%; Score 60.6; DB 6; Length 11422;
XX	Best Local Similarity 48.0%; Pred. No. 0.097;
XX	Matches 272; Conservative 0; Mismatches 284; Indels 11; Gaps 3;
OY	6 TTTAAATTTGGGAAAAATTAATAATATATGATTAATAATGCGCAATAAAGTTAGATGAGA 65
DB	7207 TTTAAATTTAGTAGGAAATATTAATAGATATTTTAAATTTGAATTAATAAATATATTAATAA 7266
OY	66 GTTTTTTTAAATTTTATATATAAATTAATTTTTCATCGAAATGTTTTTAACATATAA 125
DB	7267 GATTGTATTTTATTTTATTTTAAAGTAATGTTATATATGTTAGATATTT---TTAAA 7322
OY	126 TTTTGACATACCAACAGATATATTTCCGCCACAGATGAGCTACGAGAGTGCCTGTC 165
DB	7323 AATGTGTAGTAATTTTATTAATTTTAAATTAATAAATTTAGTTTATTTTAAATTTAAATTT 7382
OY	186 TTTAAACCTCGTAAGTTTGGTCTTACCCACCAATACCCACAAAGTAAACGACAT 245
DB	7383 TTTAAATTTAAATTTTAAATTTTAAATTTGAATGTTGAAATTTTAAAT---AAAAATTTT 7438
OY	246 ACCGGTTAATAATTTCTAACCGGTTTAAAGTTTACATATAATCATTTACTAATCCGCGT 305
DB	7439 TGTATTTATTTTAAATTTTAAATTAATTAATTTTGAATGAATGAATTTAAATTTAGATAT 7498
OY	306 GAAATTAAGTTTATGAAATGTGGTATTTTGTAGGTACAGTGAATTTATTTATTTT 365
DB	7499 ATAGTGTATTTTGGAAAAATGCTGTTTATTTTATGATAATATGTTTATGATATATGTTTT 7558
OY	366 AGTACTGTGTTTTCTTTTTTGGGTCAACTAGTACTTTTTCTTGGACATCAAAATTA 425
DB	7559 TTTTGAATTTATGTAGAGATAGATATTTTAATGTATTAGAATTT---ATTAAAGAAAACGA 7615
OY	426 TTGTAGACGAGTGTCCATATATATAGTGTGAAATGAAATGAATATTTAGTAAATTAATA 485
DB	7616 TTTAAATTTGTACGTTGTGTATATGATATTTTAAATTTAAAGTTTAAATTAATAATAATTA 7675
OY	486 AATATGAAAGTGAACAAAAAATTTAGTGGAAAAAGTAACATGGAAGAAAGAAAGCTTAT 545
DB	7676 AAAAAAAGAAAAACGATTAATGTTTGTGTAATTTTAAATTAATTAATAAATTTTGAGA 7735
OY	546 GCTTGGCAAGATTCATTAAGTTATTT 572
DB	7736 TATAGATTTTAATTTTATTTTATTTT 7762
XX	
XX	RESULT 7
XX	ABL32218
XX	ID ABL32218 standard; DNA; 11422 BP.
XX	ABL32218;
XX	26-MAR-2002 (first entry)
XX	Human immune system associated gene SEQ ID NO: 191.
XX	Human; immune system disease; cytosine methylation; antiaethmatic;
XX	antiathectoclastic; antiaethmatic; cytosine; neutrotropic;
XX	neutrotropic; anti-HIV; anticonvulsant; ophthalmological;
XX	antitubercular; antitubercular; antidiabetic; antiparasitic;
XX	antitubercular; cancer; eye disease; arteriosclerosis; anaemia;
XX	acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
XX	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX	ds.



CC Present sequence represents a specifically claimed EST isolated from a  
CC cotton variety DP50B primed seed cDNA library (LIB3825). The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from the US patent office at  
CC [seqdata.uspto.gov/sequence.html?DocID=US20040123340](http://seqdata.uspto.gov/sequence.html?DocID=US20040123340)  
XX

SO Sequence 560 BP; 268 A; 7 C; 48 G; 237 T; 0 U; 0 Other;

Query Match 8.6%; Score 60.4; DB 13; Length 560;  
Best Local Similarity 50.7%; Pred. No. 0.097;  
Matches 145; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 251 TTAATTAATTTCTACCGGTTATTAAGTTTACATAATCATTTACTTAATCCGCTGAAT 310
DB 371 TTTTCTTTTAAACACCTTTTCTTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 311 TAAAGTTTAAAGTGTGTTATTTGTAGTCAAGTGAATTTATTAATTTTGTATC 370
DB 311 TTTTCTTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 371 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
DB 251 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 431 GACGAGTGTCCATATATAGATGCGTGAATGAATGAATATGATATTAATTAATATA 490
DB 191 AATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 491 GAAAGTGAACAAAAAATTAAGTGAAGAAAGTACTGGAAGAAAGAA 536
DB 131 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 86
```

## RESULT 9

ID ABV60941/C  
ABV60941 standard; cDNA; 626 BP.

AC ABV60941;

DT 13-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 60932.

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US005171.

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-021314P.

PR 18-JUL-2000; 2000US-0215007P.

PR 13-DEC-2000; 2000US-0255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer.  
XX Claim 1; Page 11585; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX

SO Sequence 626 BP; 356 A; 15 C; 41 G; 130 T; 0 U; 84 Other;

Query Match 8.5%; Score 60.2; DB 5; Length 626;  
Best Local Similarity 48.0%; Pred. No. 0.11;  
Matches 134; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
QY 258 TATCTAACCGGTTTAAAGTTTACATTAATCATTTACTTAATCCGCTGAATTAAGTTT 317
DB 330 TNNCCNNCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 318 TATGAAGTGTGTTATTTGTAGTCAAGTGAATTTATTAATTTTGTAGTCTGTTT 377
DB 270 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 378 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
DB 210 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 438 GGTCCATATATAGATGCGTGAATGAATGAATATGATATTAATTAATGAAGTGC 497
DB 150 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
QY 498 AACAAAAAATTAAGTGAAGAAAGTACTGGAAGAAAGAA 536
DB 90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 52
```

## RESULT 10

ID ACN46935  
ACN46935 standard; cDNA; 499 BP.

AC ACN46935;

DT 02-DEC-2004 (first entry)

DE Cotton primed seed EST Clone ID: LIB3825-005-Q1-N6-D10, SEQ:1716.

XX

KM Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;

KW variety DP50B; library LIB3825; molecular tag; molecular marker;

KW genetic mapping; molecular mapping; seed germination; plant growth;

KW plant quality; plant yield; plant breeding; tissue printing; ss.

OS Gossypium hirsutum.

PN US2004123340-A1.

PD 24-JUN-2004.

PF 12-DEC-2001; 2001US-00021323.

PR 14-DEC-2000; 2000US-0255619P.

PA (DEIK/) DEIKMAN J.

PA (FENG/) FENG P C C.

PA (FING/) FINCHER K L.

PA (ZIG/) ZIEGLER T E.

XX Delkman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its  
PT fragment, useful for isolating a variety of agronomically significant  
PT genes associated with plant growth, quality or yield, and as molecular  
PT tags to map genes.

PS Claim 1; SEQ ID NO 1716; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs);  
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated  
CC from primed or non-primed seeds from variety DP50B, mature seeds from  
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium  
CC tissue, developing fibres, carpel walls and septa from variety  
CC Nuotcon33B. The invention also relates to substantially purified  
CC the proteins or their fragments encoded by nucleic acid molecules of the  
CC invention, and to transformed plants having a nucleic acid construct  
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as  
CC molecular tags to isolate genetic regions, to isolate genes, to map  
CC genes, to determine gene function and to determining whether genes are  
CC members of a particular gene family. The nucleic acid molecules may be  
CC used for isolating a variety of agronomically significant genes  
CC associated with plant growth, quality, yield, and could also serve as  
CC links in metabolic and catabolic pathways. The nucleic acid molecules are  
CC also useful for identifying genes important in initiating and maintaining  
CC seed germination or that may be used to mitigate stresses encountered  
CC during seed germination. The ESTs additionally enable the acquisition of  
CC promoters and cis-regulatory elements which will be useful to express  
CC agronomically significant genes in these tissues and/or other tissues,  
CC and also permits the acquisition of molecular markers useful in breeding  
CC schemes, genetic and molecular mapping, and in cloning of agronomically  
CC significant genes. The nucleic acid molecules are further useful for  
CC detecting the presence or quantity of a protein by tissue printing. The  
CC present sequence represents a specifically claimed EST isolated from a  
CC cotton variety DP50B primed seed cDNA library (U183825). The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from the US patent office at  
CC seqdata.uspto.gov/sequence.html?DocID=US20040123340  
XX

XX Sequence 499 BP; 212 A; 32 C; 62 G; 193 T; 0 U; 0 Other;

XX Query Match 8.5%; Score 60; DB 13; Length 499;

XX Best Local Similarity 50.7%; Pred. No. 0.11;

XX Matches 144; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

XX 305 TGTATTAAGTTTGAAGTGGTATTTGTAGTCACGGAATTTATTTATTTT 364

XX 34 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 93

XX 365 TAGTACTGTTTCTTTTGGGTTCACTAGTACTTTTCTTGACATCAAAAT 424

XX 94 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 153

XX 425 ATTGTAGACGAGTGTCCATATATAGATGTAATGAATATATATATATATA 484

XX 154 AAAAAAGGGGGTAAAAAATTTTAAAAAGAAAAAAGAGGGGGGTTAAAAA 213

XX 485 AATTATGAAGATGACAAAAAATTAAGTGAAGAAAGTAACTGGAAGAAAGCTTAT 544

XX 214 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 273

XX 545 GCGTTGGCAAGATTCCATTAAGTTTATTTCCAAAAAGGAAAGA 588

XX 274 AAAAAAGGAAAAAAGGAGGTTTAAAAAAGAAAAAAGAAAAAAG 317

XX RESULT 11

XX AB210100 standard; DNA; 8056 BP.

XX AC AB210100;

XX DT 16-JAN-2003 (first entry)

XX Haematopoietic cell proliferation disorder related DNA sequence #240.  
DE Human, haematopoietic cell proliferation disorder; cytostatic;  
XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
KW Cytosine methylation state; gene; ds.  
XX

XX Homo sapiens.

XX WO200277272-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-EP003401.

XX 26-MAR-2001; 2001US-0278333P.

XX (EPIG-) EPIGENOMICS AG.

XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Iesche R, Ieu E,  
PI Lewin A, Lipicher E, Walter S, Model F, Mueller V, Otto T, Pelet C;  
PI Schwabe I, Ziebarth H;

XX WPI; 2003-018942/01.

XX Detecting and differentiating between haematopoietic cell proliferative  
PT disorders, comprises contacting a target nucleic acid with a reagent that  
PT distinguishes between methylated and non-methylated CpG dinucleotides.  
XX

PS Claim 28; SEQ ID NO 240; 117pp; English.

XX The present invention describes a method for detecting and  
XX differentiating between haematopoietic cell proliferative disorders  
XX associated with at least 1 gene and/or their regulatory regions in a  
XX subject. The method comprises contacting a target nucleic acid in a  
XX biological sample obtained from the subject with at least 1 reagent,  
XX which distinguishes between methylated and non-methylated CpG  
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118  
XX represent specifically claimed nucleotide sequences from the present  
XX invention. Oligonucleotides from the present invention can be used: for  
XX differentiating between healthy haematopoietic cells and proliferative  
XX disorder haematopoietic cells; for differentiating between acute  
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
XX determining the cytosine methylation state and/or single nucleotide  
XX polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
XX related sequences and their complements; and as primers for the  
XX amplification of haematopoietic cell proliferation disorder related DNA  
XX sequences. The nucleotide sequences from the present invention can also  
XX be used for detecting a predisposition to, differentiation between  
XX subclasses, diagnosis, prognosis, treatment and/or monitoring of  
XX haematopoietic cell proliferative disorders. The present method enables a  
XX highly specific classification of haematopoietic cell proliferative  
XX disorders allowing for improved and informed treatment of patients

XX Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

XX Query Match 8.4%; Score 59.4; DB 8; Length 8056;

XX Best Local Similarity 44.7%; Pred. No. 0.16;

XX Matches 272; Conservative 0; Mismatches 336; Indels 1; Gaps 1;

XX 11 ATTGGGAAAAATTTATATATATATATATATATATATATATATATATATATAT 70

XX 2109 ATTAT 2168

XX 71 TTTTAAATTTTAT 130

XX 2169 ATTAT 2228

XX 131 ACATTAACCAAGTATATATATATATATATATATATATATATATATATATAT 190

XX 2229 AATTTAAT 2288



PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI: 2002-154758/20.  
DR  
XX Nucleic acid, useful for diagnosis and therapy of diseases associated  
PT with cell signalling e.g. cancer, comprises chemically modified genomic  
PR sequences of genes associated with cell signalling.  
XX  
XX Claim 1, SEQ ID NO 358; 24pp + Sequence Listing; English.  
XX  
XX The invention relates to a nucleic acid comprising a sequence of at least  
CC 18 bases of a segment of chemically pretreated DNA of genes associated  
CC with cell signalling. The activity of the modified sequences of the  
CC invention may be described as cytostatic. The object of the invention is  
CC to provide the chemically modified DNA of genes associated with cell  
CC signalling, as well as oligonucleotides and/or PNA-oligonucleotides for  
CC detecting cytosine methylations, as well as a method which is  
CC particularly suitable for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with cell signalling. The  
CC chemically modified DNA provided by the invention is useful for diagnosis  
CC and therapy of diseases such as solid tumours and cancer. The sequences  
CC given in records AB170111-AB170626 represent chemically pre-treated  
CC genomic DNA's of genes associated with cell signalling. Note: The  
CC sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office  
XX  
SQ Sequence 15954 BP; 4759 A; 139 C; 3187 G; 7869 T; 0 U; 0 Other;  
Query Match 8.4%; Score 59.4; DB 6; Length 15954;  
Best Local Similarity 51.4%; Pred. No. 0.16;  
Matches 162; Conservative 0; Mismatches 151; Indels 2; Gaps 1;  
QY 230 AAGGGAACGACGACCGTATTAATATCTACCGTTATTAAGTTTACATAAATC 289  
Db 11745 AATTAATTAATGTTTATTAAGTTAATTAATTAATTAATTAATTAATTAAT 11804  
QY 290 ATTACTAATCCGCGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 349  
Db 11805 ATTAGATTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11864  
QY 350 AATTATTAATTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 409  
Db 11865 AATGATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11924  
QY 410 TTGACATCAAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 469  
Db 11925 TTGAGATT--AATGAATCGAATTAATTAATTAATTAATTAATTAATTAATTAAT 11982  
QY 470 ATTGAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 529  
Db 11983 TTAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12042  
QY 530 AAGAAAAAGCTTATTT 544  
Db 12043 TAATATTATTATTATTT 12057  
RESULT 14  
ADQ24856/c  
ID ADQ24856 standard; DNA; 6816 BP.  
XX  
XX ADQ24856;  
XX  
DT 26-ANG-2004 (first entry)  
XX  
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 7676.  
XX  
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KM de.  
XX

OS Homo sapiens.  
XX  
XX WO2004048938-A2.  
XX  
XX  
XX 10-JUN-2004.  
XX  
XX  
XX 26-NOV-2003; 2003WO-US038193.  
XX  
XX 26-NOV-2002; 2002US-0429739P.  
XX  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX  
XX Aziz N, Ginsburg MM, Zlocnik A;  
PI WPI: 2004-441208/41.  
DR  
XX  
XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.  
XX  
XX Example 2, SEQ ID NO 7676; 210pp; English.  
XX  
XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.  
XX  
SQ Sequence 6816 BP; 2121 A; 1327 C; 1578 G; 1571 T; 0 U; 219 Other;  
Query Match 8.4%; Score 59.2; DB 12; Length 6816;  
Best Local Similarity 44.3%; Pred. No. 0.17;  
Matches 186; Conservative 0; Mismatches 222; Indels 12; Gaps 2;  
QY 5 TTTTAAATTTGGGAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 64  
Db 6653 TTTATATNNNANTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6594  
QY 65 AGTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 124  
Db 6593 TTTTATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6534  
QY 125 ATTTGACATTAACCAAGTATTAATTTCCGCCAGATTAAGTCTACGAGAGATGCTCGT 184  
Db 6533 TTTTATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 6474  
QY 185 CTTTAACTCGTAAGTTTGTGCTTACCAACCAATACCAACCAAGTAACGACCA 244  
Db 6473 AAATAAANTTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTA 6419  
QY 245 TACCGTTATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 304  
Db 6418 NNNANTTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 6364  
QY 305 TGTAAATTAAGTTTATGAATGCTGTTATTTGTAGTCAACGGAATTAATTAATTTT 364  
Db 6363 --TAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTNT 6306  
QY 365 TAGTACTGTTTCTTTTGTGGTCAACGTACTTTTCTTCTTGACATCAAAAT 424  
Db 6305 NTTTTNNNNNNNNNTTTTNTTTTNNNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTNTTT 6246  
RESULT 15



ABLS5644  
 ID ABL5644 standard; DNA; 50000 BP.  
 XX  
 AC ABL5644;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE AMEPV genome fragment#2.  
 XX  
 KM AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;  
 XX genetic deficiency disorder; ds.  
 OS Amsacta moorei entomopoxvirus.  
 PN WO200212526-A2.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 10-AUG-2001; 2001WO-US025287.  
 XX  
 PR 10-AUG-2000; 2000US-0224479P.  
 XX 14-SEP-2000; 2000US-00662254.  
 XX  
 PA (UYFL ) UNTV FLORIDA.  
 XX  
 PI Moyer RW, Li Y, Bawden AL;  
 DR WPI; 2002-227161/28.  
 XX  
 PT Novel recombinant entomopox virus vector useful for delivering  
 PT polynucleotide encoding protein to vertebrate cell, comprises  
 PT polynucleotide encoding protein operably linked with heterologous  
 PT promoter sequence.  
 XX  
 PS Disclosure; Page 150-175; 326pp; English.  
 XX  
 XX The invention relates to a recombinant entomopox virus (EPV) vector,  
 CC comprising a polynucleotide encoding a protein operably linked with a  
 CC heterologous promoter sequence. The invention also concerns methods for  
 CC providing gene therapy for genetic deficiency disorders. Vectors of the  
 CC invention are useful for delivering a polynucleotide encoding a protein  
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.  
 CC The vector is introduced into the vertebrate cell by infection in a viral  
 CC particle, or by transfection, transduction, or injection either in vitro  
 CC or in vivo. The vector is useful for the delivery and expression of  
 CC biologically useful proteins in gene therapy protocols, and for  
 CC delivering large DNA segments for engineering of vertebrate cells.  
 CC Polynucleotides of the invention have applications in techniques such as  
 CC their use as insertion sites for foreign genes of interest, hybridisation  
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the  
 CC production of sense or antisense nucleic acids. Vectors of the invention  
 CC provide for stable integration and expression of heterologous DNA in host  
 CC cells, and are adapted for accepting large heterologous polynucleotide  
 CC inserts which can be delivered in an infected or transformed cell and  
 CC expressed in a stable fraction. The current sequence represents a  
 CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei  
 CC (AMEPV)  
 CC  
 SQ Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 U; 0 Other;  
 Query Match 8.4%; Score 59.2; DB 6; Length 50000;  
 Best Local Similarity 54.8%; Pred. No. 0.18;  
 Matches 138; Conservative 0; Mismatches 113; Indels 1; Gaps 1;  
 QY 260 TTCTAACCGGTTTATAGTTTACATTAATTCGCGTAAATTAAGTTTA 319  
 DB 49095 TTATTAATCTTATTTTATTAATTTATCAATTTTATATATATTTATTCAAATTAAT 49154  
 QY 320 TGAATAAGGTTTATTTAGGTCAGGTAAATTTATTAATTTTATAGTCTTTTC 379  
 DB 49155 TAAATTTTATTTTATTTATTAATTAAGATTTTATTAATTTTGTATGTTATTAAT 49214  
 QY 380 TTTTGGGTCACTAGTACTTTTCTTTCATCAATCAAAATTAATTTAGTCAGGTGG 439

DB 49215 AAATTTTAAATATATATATATATATTTGTTTAAATTTTAAATCAATTAATTTTAAACATGCCG 49274  
 QY 440 TCCATATATAGATGCTGGAATGAAATGAATATTTGAGTAAATTAATTAATATAGAAAGTGA 499  
 DB 49275 AAATTAATATATATATGTTAAATATTTTATTAATTAATTAATTTTCAATTAATTAATTTGTT 49333  
 QY 500 CAAAAAAATTA 511  
 DB 49334 CTAATATATTA 49345

Search completed: November 17, 2005, 15:59:37  
 Job time : 479 secs

*This Page Blank (uspto)*







```

Db      324 ATMAATAMAAATTAATTTTWTWTWTWTATTTTATTTTWTATTTATATMAAT 383
Qy      301 CGGGTAAATTAAGTTTATGAATGGTATTTTGTAGTCACGGAATTTATTAAT 360
Db      384 MATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 443
Qy      361 TTTTATGACTGTTTCTTTTGGTTCACATGATTTTCTT 412
Db      444 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 495

RESULT 5
CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR08K10 of RCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921 GI:4941778
VERSION AL063921
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
18ogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACR08K10"
/clone_1ib="RCI-98"
/note="end : TET3"

ORIGIN
Query Match 10.3%; Score 72.4; DB 9; Length 1101;
Best Local Similarity 18.5%; Pred. No.0.0039;
Matches 125; Conservative 283; Mismatches 264; Indels 2; Gaps 2;

Qy 30 TATATGATTAATGCAATAAAGTTAGATGAGATTTTAAATTTTATATTA 89
Db 1100 KARRMDDDMRTRDRDMDMTKMTWMDRADMAGDADRWAMDDGATMTATMM 1041
Qy 90 AATAATTTTTCATGAATGTTTATGATTAATTTGACATTAACAAGTATTTT 149
Db 1040 MMTTATTTTTCATGAATGTTTATGATTAATTTGACATTAACAAGTATTTT 191
Qy 150 TCCGACCAATGAGCTAGAGAGATGCTGCTTTAAACCTGTAAGTTTGTG 209
Db 980 GRRDGRKRRDKRRDGDGDKGKKKAKAKAKAKATKMTWMDWMDKMKMGADORKA 921

```

```

Qy 210 TTACCAACCCATATCCCAAGTAACGACATACCGGTTATTAATTTTACCGG 269
Db 920 DDDGAGKDDDDGKDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD 861
Qy 270 TTTATGATTTTATTAATGATTTTATTAATGATTTTATTAATGATTTTATTA 328
Db 860 WADADMTWDAADDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD 801
Qy 329 GTTATTTTGTAGTCACGTAATTTTATTAATTTTATTAATTTTATTAATTT 388
Db 800 RRRADDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD 741
Qy 389 TTCACTAGTACTTTTCTTTGACATCAAAATTTTGTAGACAGCGTCCATAT 448
Db 740 AGTAGKRRRTWRKRRRDRDTRDADADTRDTRDTRDTRDTRDTRDTRDTR 681
Qy 449 AGATGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 508
Db 680 RATWDRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTR 621
Qy 509 TTATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 568
Db 620 KMDWTRADRRMDRRMDRRMDRRMDRRMDRRMDRRMDRRMDRRMDRRMD 562
Qy 569 ATTCAACCAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 628
Db 561 ATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 502
Qy 629 TGTAAAGTAGAGAGCTCAATTTCTAGTTCATCAAAATTTGCGTTATTTCT 688
Db 501 MAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 442
Qy 689 GTGTCACCTGAAT 702
Db 441 TTTTWTWTWMAAT 428

RESULT 6
CNS010MP/c 734 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN04I20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL099163 GI:5610774
VERSION AL099163
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 734)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.biol.ac.uk-. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billard at CEH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source location/Qualifiers
1..734
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACN04I20"
/clone_1ib="DrosBAC"
/plasmid="pBelOBAC11"

```

```
/note="end : T7"
```

QY	25	AATAAATATATGTTTAAATGCGCAAAATTAAGTAATAGAGAGTTTAAAAATTTTAT	84
Db	634	AATGAATTTTAAATGTTAAAAAAGAAAAAAMGATGATMAATAGTATTTWTAATGTTKGTDT	575
QY	85	TATPAAATATATTTTTGTCATGAAATTTGTTTTAAATPAAAAATTTTGACATTAACCAATAT	144
Db	574	TTSATTPAAAAATMDGGTAAATGOMGTAATTTAAATATGATGAKATTTAAAKGMATBK	515
QY	145	TATTTTTCCGCCAGATTTAGCTCTACGAGAAATGTCCTGTTTAACTCGTAAAGTTT	204

Db	514	GRKTTTGATTAATAATKTTTTTTTDTTGGGGGAATTTKTAIRNRNTAATTTAABAAAAAT	455
QY	205	TGCTCTTACCCAAACCAATACCCACAAAGTAACGACCATACCGGTTAATAATATTCTA	264
Db	454	GGGTGGAAAGTATTAAAAATATAAGAAWTAATAAATAATTTGGWTTTTTAAGTGT	359
QY	265	ACCGGTTTAAAGTTATACATAATCATTTACTTAATCCCGGTGATTAAGTTATAGAA	324
Db	394	ATTAAMWTAKMAWTTTGTAATKMTTTTAAGTCTTTTWTATTAATMTWTTWATTT	335
QY	325	TGTGTTATTTTGGAGTCAAGTGAATTTAATTAATTTTTTAACTGTTGTTTTCTTTT	384
Db	334	WTTTTTMMTTATATWTTTTTGATTTMTTTTDDGKMAAATTAATAATGWTMTTKATTTT	275
QY	385	TGGGTCACATPAGTACTTTTCCCTTGGACATCAAAATTTGTAGACAGAGTGTCAT	444
Db	274	TGTGTGMTTTTWTKTAATAMWMMMKTTTTTKMAATTTKTTTAAATTAATAAT	215
QY	445	ATATAGATGCGTGAATG-----AAATGAATATTGAGTAATPAAATAATATAGAAC	495
Db	214	KWTWTWTWTTTTMMKMTWTATWAMAAAAATWTBKKMMWTTTAAAAAAATTTTTAAT	155
QY	496	TGAACAAAAAATAATAGTGAAGAAAGCTAACGGAAGAAGAAAGGCTTATTCGCTTGGCAG	555
Db	154	KAAAAAATATTTTDXAAAAAKAAAAATAAATATTTTDXMTAMWTTTWTTTTTTCKAAAT	95
QY	556	ATTCCATAAGTTTATTTACCAAAAAAGAAAGAGTACTTGCCCTTCTCTCTCTTT	615
Db	94	TGGAAMKTTTWTATWATWTATATATAKATATATAATATATKTWTTTTTATAAATTTT	35
QY	616	TAAATGCTAACTG 630	
Db	34	TTAKGKNNNAATATK 20	

RESULT 7	
CNS016CO	
LOCUS	CNS016CO 1200 bp DNA linear GSS: 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106578
VERSION	AL106578.1 GI:5622626
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1200)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 101 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a

Location/Qualifiers  
1. .1200

collaboration with the European Drosophila Genome Project (EDGP) <http://www.edgp.ebi.ac.uk>. This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billand at CSH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC1.

```
/note="end" : T7"
```

[illegible]

Db	881	TTTTTATMAAAAATADGANTATATTTWACGAAAAATTTTAAWAAATTTATTTWWAAATAAW	940
Qy	371	TTG-TTTTCTTTTGGGTTCAACTAGTACTTTTTCTTGACATCAAAATTAATGT	429
Db	941	AAAAATTTTTTAAATTTAAATWWTATTTTAAATTTTAAWAAAAAATTTTTTTTTAAWAT	1000
Qy	430	AGACGAGTGGTCATATATAGATGTAATGAATGAATATATGAGTAAATAAATATAT	489
Db	1001	AAWMTAAWAAATMAAAWMTATTTWAAATTTAAATAATTAATTTWAAWAAWMTWMAAAWAAA	1066
Qy	490	AGAAAGTGAACAAAAAATTAATGTGAAAAAGTAACTGAAAAAGAAAGCCTTAATGGCTT	549
Db	1061	AAWMTAAWMAAAAAAAMAAAAAAMAAAAAAMAAAAAAMWMTAATTAWMAWMAWMAWMAATTT	1120
Qy	550	GGCAAGATTCATAGTTTATTTTCCACCAAAAGGAAAGAGTACTGGCCTTCTCTCT	609
Db	1121	WTAAAAWMTAAWMTAAWMTTTTTTTTMTWMAWATTTAATTTATTAWMTWMAATTAWMAATAAW	1180
Qy	610	CTCTTTTAAAT 621	
Db	1181	ATATAATATAW 1192	

RESULT 8					
LOCUS	CR722884				
DEFINITION	CR722884	1896 bp	mRNA	linear	HTC 19-AUG-2004
ACCESSION	CR722884				
VERSION	CR722884.1	GI:51221135			
KEYWORDS	HTC; cDNA; full-length; Tetraodon nigroviridis.				
SOURCE	Tetraodon nigroviridis				
ORGANISM	Tetraodon nigroviridis				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Acanthoidei; Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 1896)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-AUG-2004)	Genoscope - Centre National de Sequencage -			
	2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE				











**Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tübingen D-72076, Germany**  
**Tel.: 00487071601371**  
**Fax: 00487071601498**  
**Email: ralf.sommer@cuebingen.mpg.de**  
**Class: BAC ends.**

**FEATURES**  
**SOURCE**

/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

Search completed: November 17, 2005, 17:46:28  
Job time : 3124 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 17, 2005, 15:35:13 ; Search time 173 Seconds  
(without alignments)  
6668.064 Million cell updates/sec

Title: US-10-810-788a-4

Perfect score: 705

Sequence: 1 aagcttcaattcgga...ctctgctcactgaatacc 705

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database: Issued Patents\_NA:\*

- 1: /cgn2\_6/ptodaca/1/ina/5A COMB.seq:\*
- 2: /cgn2\_6/ptodaca/1/ina/5B COMB.seq:\*
- 3: /cgn2\_6/ptodaca/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodaca/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodaca/1/ina/PTCUS COMB.seq:\*
- 6: /cgn2\_6/ptodaca/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	53.2	7.5	1044	US-09-270-767-13657	Sequence 13657, A
C 2	51.6	7.3	19124	US-08-487-826B-13	Sequence 13, Appl
C 3	50.8	7.2	187169	US-09-949-016-12776	Sequence 12776, A
C 4	50.8	7.2	191569	US-09-949-016-15940	Sequence 15940, A
C 5	50.6	7.2	640681	US-09-980-988-1	Sequence 1, Appl
C 6	50	7.1	1141	US-09-806-708B-22	Sequence 22, Appl
C 7	49.8	7.1	192302	US-09-949-016-15270	Sequence 15270, A
C 8	49.6	7.0	636	US-08-998-416-1137	Sequence 1137, Ap
C 9	49.6	7.0	1141	US-09-806-708B-22	Sequence 22, Appl
C 10	49.4	7.0	96922	US-09-949-016-17061	Sequence 17061, A
C 11	49.2	7.0	601	US-09-949-016-124813	Sequence 124813, A
C 12	49.2	7.0	6996	US-09-949-016-28913	Sequence 28913, A
C 13	48.8	6.9	601	US-09-949-016-28913	Sequence 28913, A
C 14	48.8	6.9	601	US-09-949-016-125860	Sequence 125860, A
C 15	47.8	6.8	927	US-09-270-767-29671	Sequence 29671, A
C 16	47.8	6.8	601	US-09-949-016-28912	Sequence 28912, A
C 17	47.6	6.8	601	US-09-949-016-28912	Sequence 28912, A
C 18	47.6	6.8	601	US-09-949-016-125859	Sequence 125859, A
C 19	47.6	6.8	601	US-09-949-016-125861	Sequence 125861, A
C 20	47.6	6.8	144322	US-09-949-016-15316	Sequence 15316, A
C 21	47.6	6.8	194537	US-09-949-016-12928	Sequence 12928, A
C 22	47.6	6.8	201529	US-09-949-016-12740	Sequence 12740, A
C 23	47.4	6.7	1055	US-09-806-708B-23	Sequence 23, Appl
C 24	47.4	6.7	95561	US-09-949-016-12768	Sequence 12768, A
C 25	47.4	6.7	95561	US-09-949-016-13306	Sequence 13306, A
C 26	47.4	6.7	95561	US-09-949-016-13307	Sequence 13307, A
C 27	46.6	6.6	601	US-09-949-016-109980	Sequence 109980, A

C 28	46.6	6.6	601	US-09-949-016-110061	Sequence 110061, A
C 29	46.6	6.6	113060	US-09-949-016-14773	Sequence 14773, A
C 30	46.6	6.6	113060	US-09-949-016-14774	Sequence 14774, A
C 31	46.4	6.6	601	US-09-949-016-30531	Sequence 30531, A
C 32	46.4	6.6	601	US-09-949-016-37150	Sequence 37150, A
C 33	46.4	6.6	601	US-09-949-016-37164	Sequence 37164, A
C 34	46.4	6.6	601	US-09-949-016-145868	Sequence 145868, A
C 35	46.4	6.6	601	US-09-949-016-146136	Sequence 146136, A
C 36	46.4	6.6	601	US-09-949-016-146404	Sequence 146404, A
C 37	46.4	6.6	1784	US-09-601-198-21	Sequence 21, Appl
C 38	46.4	6.6	4818	US-08-817-926-27	Sequence 27, Appl
C 39	46.4	6.6	79835	US-09-949-016-12456	Sequence 12456, A
C 40	46.4	6.6	79835	US-09-949-016-16121	Sequence 16121, A
C 41	46.4	6.6	205044	US-09-949-016-15851	Sequence 15851, A
C 42	46.4	6.6	205044	US-09-949-016-15852	Sequence 15852, A
C 43	46.4	6.6	205044	US-09-949-016-15853	Sequence 15853, A
C 44	46.4	6.6	223471	US-09-949-016-12387	Sequence 12387, A
C 45	46.4	6.6	223471	US-09-949-016-12724	Sequence 12724, A

## ALIGNMENTS

```
RESULT 1
US-09-270-767-13657/c
; Sequence 13657, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OR INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 13657
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-13657

Query Match      7.5%; Score 53.2; DB 4; Length 1044;
Best Local Similarity 53.3%; Pred. No. 0.039;
Matches 112; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

OY 21 ATTAAATATATATGATTAATGCAATTAAGTAAAGTAAATTTTAAATTT 80
    |||||
DB 320 AATTAATTAAGATTTGTTCAATTAACATCGATTTTGCTTGAGTTTATCTTTT 261
    |||||
OY 81 TTAATTAATTAATTAATTTTGCATGAATGTTTAAAGTAAATTTGACATTAACCA 140
    |||||
DB 260 CTCCTTGATCTTTTGTATGTATCTTAAGCTTTAGCTTAATTTTTCGGAAGTAT 201
    |||||
OY 141 GATTAATTTTCCGCAAGAAATGATGATGAGATGCTCTTTAAACCTCGTAA 200
    |||||
DB 200 CTTTGTATTAATGACATCGCACTTAATTAAGCTGACCAACTATTAACCTTAACA 141
    |||||
OY 201 GTTTGCTTACCAACCAATTAACCA 230
    |||||
DB 140 CTTTGTATTAATGACCAACCAATTAACCA 111
    |||||

RESULT 2
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Willems, Thomas E.
```





```
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1141)
OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22
```

```
Query Match
Best Local Similarity 7.1%; Score 50; DB 4; Length 1141;
Matches 74; Conservative 226; Mismatches 336; Indels 4; Gaps 1;
```

```
QY 49 ATAAAGTTAGAGAGAGTTTAAATTTTAAATTAATTAATTTTTCATGAAA 108
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 936 WTTDVRMAKAKANNNNNNNAATACYNRAATNNMMATHMKHGAASKRIHHTCR 877
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 109 TTGTTTTAAGATAAATTTTGACATACCAAGTATTTTCCGACCAATGAGTC 168
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 876 RTYNNNNNNNARIVYVYHHAARRMNAWRTNNNNNNNNNACNRTMTWMAWKSQCN 817
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 169 ACGAGAGATGCTGCTCTTTAAACCTCGTAAGTTTGTCTTACCCCAATACCA 228
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 816 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 757
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 229 CAAAGGTAACGACGATCCGTTAATATTTCTAAGCTTATGAGTTTCAATAA 288
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 756 TMTNTTMYDMWTTTBTBTTRNTTSTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 697
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 289 CATTACTAATCCGCGTGAATTAAGTTTATGAAATGCTTATTTTGTAGTCACGTG 348
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 696 CWMNNTDARTNTTNTVWRRTMTNTKTRWSTTRBHYGATNNNNNNNNNNNNNNNSC 637
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 349 AAATTTAATTTTATTTAGTCTGTTTCTTTTGGG---TTCACCTAGTACTTT 404
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 636 CTCTRMWTKWTKGDMTRKKVKWRDITCTYVDWADSWVWYANWMCSDVYTYRNN 577
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 405 TTTCCTTGACATCAAAATTTATGAGAGTGGCCCATATAGATGCTGAAGAAA 464
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 576 TYKSAHSTWYWSNNNAWTRYSARWWSMAATTTNNNNWMSGBVWRAGTMMRHNN 517
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 465 TGAATATGAGTAAATTAATATAGAAAGTGAACAAAAAATTAAGTGAAGTAA 524
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 516 NNNDDTTRYWMMWRMBTTTYVDSMCKAKSMWRGNWRRAMKMMWANNDAAGMHTY 457
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 525 CTGGAAGAAGAAAGCTTATTTGCTTGCAAGATTCATTAAGTTTACCAAAAAAGA 564
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 456 MGNNTWMMBRAMKMMWMCRAAYCCNNNNNNRACVWHKHMRTWKYMMKAAACNNNNBK 397
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 585 AAGAGACTTGCGCTGCTCTCTCTCTTTAAATGTAACGTGTAAGTAGAGA 644
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 396 AMTRVAMMYSRDNTTMDMMMTSDMBHWYTVDTMMERAMNNNNNNNNNNNNNNNN 337
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 645 CTCATTTCTAGTTCATCAAAACATATTTGCGCTTATTAAT 684
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 336 WMDHWNTHCTGNNNTGSAIYBMAAAGMMAAGANBVTYNN 297
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

```
RESULT 7
US-09-949-016-15270/c
Sequence 15270, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949, 016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15270
LENGTH: 192302
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(192302)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15270
```

```
Query Match
Best Local Similarity 7.1%; Score 49.8; DB 4; Length 192302;
Matches 184; Conservative 0; Mismatches 197; Indels 6; Gaps 1;
```

```
QY 24 TATATATATGATTAATTAATGCAATTAAGTATGAGAGTTTAAATTTTAA 83
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 76211 TGAATATTTTGGTTAAATATCAGAGTAAATTTAAATATAATAGATGCAATTAAGT 76152
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 84 TTATAAATTAATTTTTCATGAATTTGTTTAAAGATTAATTTTGAATTAACCAAGTA 143
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 76151 TTTTAAATTTCTTTTAACAACGATTAATTTTCAAGTAAATAATGAGACAGTAAATAGA 76092
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 144 TTATTTCCGACAGATGAGCTACAGAGATGCTT-----GCTTTAACTCGT 197
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 76091 GAATCTACCTTTTGTGAGAGAACAACTGTTCCCACTCTAGTAAATTAATTAAGTGTG 76032
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 198 AAAGTTTGTCTTACCCCAACCAATACCACAAAGTAAAGACCAATCCGTTAATA 257
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 76031 TTTTTTTTTTTTTTTAAAGAAAGTACAGAGCTGATGATAAATCTGCTCGGAG 75972
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 258 TATTTACCGGTTTAAAGTTTACATTAATCATATTACTTAATCCGCTAATTAAGTTT 317
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 75971 TACTCTCTGCTGCTATTTGTAATGATCAATTTTCATGCTGATACCAATAT 75912
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 318 TATGAATGATGTTATTTGTAGTACGTAATTAATTAATTTTATTTTGTACTGTTT 377
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 75911 TCTAAAAGAGCTTAATTAATTTAGAGGATGAATGATTTAGTCATGTTTAGTGAATTTT 75852
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 378 TCTTTTGGGTTCACACTACTTCTT 404
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 75851 TTGTTTCAATTTAATCGACAGACATGT 75825
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

```
RESULT 8
US-08-998-416-1137
Sequence 1137, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Phillipsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
```



? FILING DATE: 24-DEC-1997  
? CLASSIFICATION: 435  
? PRIOR APPLICATION DATA:  
? APPLICATION NUMBER: CH 0016/97  
? FILING DATE: 31-DEC-1996  
? ATTORNEY/AGENT INFORMATION:  
? NAME: Meigs, J. Timothy  
? REGISTRATION NUMBER: 38, 241  
? REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGL1976  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: 919-541-8587  
? TELEFAX: 919-541-8689  
? INFORMATION FOR SEQ ID NO: 1137:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 636 base pairs  
? TYPE: nucleic acid  
? STRANDEDNESS: single  
? TOPOLOGY: linear  
? MOLECULE TYPE: DNA (genomic)  
? ORIGINAL SOURCE:  
? ORGANISM: PAGI692RP  
?  
? US-08-998-416-1137

CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 17061  
LENGTH: 96922  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17061

Query Match 7.0%; Score 49.4; DB 4; Length 96922;  
Best Local Similarity 44.9%; Pred. No. 0.45;  
Matches 188; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

QY 272 TATAGTTTACATTAATCATTTACTAATCCCGGTAAATTAAGTTTATGAAATGGTGT 331  
DB 38390 TAT 38449  
QY 332 ATTTGTAGTGCACGTGAATTTATTAATTTTGTAGTCTTTTCTTTTGGGTTTC 391  
DB 38450 TAT 38509  
QY 392 AACTAGTACTTTTCTTCTTGTACATCAAAATTTTGTAGACAGTGGCCATATATAGA 451  
DB 38510 TAAATATGTTTAT 38569  
QY 452 TCGTAAATGAATGAATTTAGTAAATTAATTAATATAGAAAGTGAACAAATTAATTA 511  
DB 38570 TGTAT 38629  
QY 512 GTGAAAAGGTACTGAAAGAAAGGCTTATTTGGCTTGGCAAGATTCATTAAGTTTAT 571  
DB 38630 TTTAT 38689  
QY 572 TCACCAAAAAGAAAGAGTACTGGCTTCTCTCTCTCTCTCTTTTAAATGTTAAACTGG 631  
DB 38690 TATTTAT 38749  
QY 632 TAAAGTAGAGACCTAATTTCTAGTTCATCAAAATTTGGCGTTATATTTCTGT 690  
DB 38750 AATAGTAAATACAT 38808

RESULT 11  
US-09-949-016-124813/C  
Sequence 124813, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 124813  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-124813

Query Match 7.0%; Score 49.2; DB 4; Length 601;  
Best Local Similarity 47.6%; Pred. No. 0.24; Mismatches 193; Indels 6; Gaps 1;  
Matches 181; Conservative 0;

QY 24 TAT 83  
DB 384 TAT 325  
QY 84 TTTAT 143  
DB 324 TTTTAAATCTTTTAAACAGAACTTTTCAAGTAAATATGACACAGTAAATAGA 265  
QY 144 TTTATTTCCCGCAAGATTTGATGCTACAGAGATGTCTT-----GTCTTAACTCGT 197  
DB 264 GAATTTTCACTTTTGTGAGAGAACACTGTTCCCACTCTAGTATATTTAAACTGTG 205  
QY 198 AAAGTTTGTCTTACCAACCAATACCAACCAAGTAAACGACATACCGGTATATTA 257  
DB 204 TTTTATTTTATTTTATTTTAAAGAAAGTACAGTGTCTGTAGTAAATATCTCGGAG 145  
QY 258 TATTTCAACCGTTTAT 317  
DB 144 TACTCTGCTGTACTAT 85  
QY 318 TATGAATGTGTATATTTTGTAGTACAGTAAATTTATATTTTATATTTTATAGTAT 377  
DB 84 TCTAAAGAGCTTGAATTTGAGGAGTAAATGATTTAGTATGCTATGTTTAAAGTAAATTT 25  
QY 378 TCTTTTGGGTCAACTAG 397  
DB 24 TGTTCATATTTAATCGAG 5

RESULT 12  
US-09-573-080A-29  
Sequence 29, Application US/09573080A  
Patent No. 6828097  
GENERAL INFORMATION:  
APPLICANT: JOAN, KNOEL  
TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATION  
FILE REFERENCE: 30307  
CURRENT APPLICATION NUMBER: US/09/573,080A  
CURRENT FILING DATE: 2000-05-16  
NUMBER OF SEQ ID NOS: 479  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO: 29  
LENGTH: 6996  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: repeat region  
LOCATION: (1)..(6996)  
OTHER INFORMATION: heryg25  
NAME/KEY: misc feature  
OTHER INFORMATION: n is a, c, g or t  
PUBLICATION INFORMATION:  
AUTHORS: Jurka, J, Malichiewicz, J, Milosavljevic, A  
TITLE: Prototypic sequences for human repetitive DNA  
JOURNAL: Journal of Molecular Evolution  
VOLUME: 35  
ISSUE: 4  
PAGES: 286-291  
DATE: 1992-10-1  
DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)  
DATABASE ENTRY DATE: 1996-01-26  
US-09-573-080A-29

Query Match 7.0%; Score 49.2; DB 4; Length 6996;  
Best Local Similarity 38.0%; Pred. No. 0.34;







Qy	61	GGAGAGTTT	TTTAAATTTT	TTTATATAAATTTT	TGACGAATGGTTTT	TAAGA	120	
Db	61	TGAGAGTTTT	TTAAATTTTT	TTATATAAAATTTTT	TGATGAATGTTTTT	TAAGA	120	
Qy	121	TAAATTTT	TGACATPAC	CCAGATATAT	TCGCGCACGAAATT	GACTACGAGATGTC	180	
Db	121	TAAATTTT	TGACATPAC	CCAGATATAT	TCGCGCACGAAATT	GACTACGAGATGTC	180	
Qy	181	CTGTCTT	TAAACCT	CGTAAAGTTT	TGGTCTTAC	CCACCAATACCCCAAAAGTTAAACG	240	
Db	181	CTGTCTT	TAAACCT	CGTAAAGTTT	TGGTCTTAC	CCACCAATACCCCAAAAGTTAAACG	240	
Qy	241	ACCATACCG	GTAAATATAT	TCTTAAACCG	GTATAAGTTTACATPAAATCATTTACTTAATC	300		
Db	241	ACCATACCG	GTAAATATAT	TCTTAAACCG	GTATAAGTTTACATPAAATCATTTACTTAATC	300		
Qy	301	CGCGGTAT	TAATTAAGTTT	TATGAAATGTGTTATTTT	TGAGTCA	CGTGAATTTTATAT	360	
Db	301	CGCGGTAT	TAATTAAGTTT	TATGAAATGTGTTATTTT	TGAGTCA	CGTGAATTTTATAT	360	
Qy	361	TTTTTAGT	ACTTGT	TTTTCTTTTT	TGGGTCAAC	TAGTACTTTTTCTTTCTTGGACATCA	420	
Db	361	TTTTTAGT	ACTTGT	TTTTCTTTTT	TGGGTCAAC	TAGTACTTTTTCTTTCTTGGACATCA	420	
Qy	421	AATATTGT	TAGACGAGT	GCATATAT	TAGATGGAATGAAATGAATATTTGAGTAATA	480		
Db	421	AATATTGT	TAGACGAGT	GCATATAT	TAGATGGAATGAAATGAATATTTGAGTAATA	480		
Qy	481	AATATAAT	TATGAAAGT	GAACAAAAAAAT	TAGTGGAAAAAGTTA	CTGAAAGAAAAGGCT	540	
Db	481	AATATAAT	TATGAAAGT	GAACAAAAAAAT	TAGTGGAAAAAGTTA	CTGAAAGAAAAGGCT	540	
Qy	541	TATTGGCT	TGGCAAGAT	CCATPAGTTTATTT	CACAAAAAGGAAGAT	CTTGGCCCT	600	
Db	541	TATTGGCT	TGGCAAGAT	CCATPAGTTTATTT	CACAAAAAGGAAGAT	CTTGGCCCT	600	
Qy	601	GCTCTCT	CTCTTTT	TAAATGTTAA	CTGTAAAGATAGAGACT	CAATTTT	TAGTTCA	660
Db	601	GCTCTCT	CTCTTTT	TAAATGTTAA	CTGTAAAGATAGAGACT	CAATTTT	TAGTTCA	660
Qy	661	TCAAAA	CAATATTT	TGGCGTTATATTT	CTGTGTGCT	CACTTGAATACC	705	
Db	661	TCAAAA	CAATATTT	TGGCGTTATATTT	CTGTGTGCT	CACTTGAATACC	705	

```

RESULT 2
US-10-810-788A-3
; Sequence 3, Application US/10810788A
; Publication No. US20040216184A1
; GENERAL INFORMATION:
; APPLICANT: Ruzekinsky, Diane
; TITLE OF INVENTION: Novel Plant Promoters for Use in Early Seed Development
; FILE REFERENCE: REN-00-118
; CURRENT APPLICATION NUMBER: US/10/810,788A
; CURRENT FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/458,828
; PRIOR FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-10-810-788A-3

```

Query Match	99.9%	Score 704;	DB 21;	Length 2017;
Best Local Similarity	100.0%	Pred. No. 7.1e-122;		
Matches 704; Conservative	0;	Mismatches	0;	Gaps 0;

QY 1 AAGCTTTAAATTTGGAAAAATTTAATAATATATGATTTAAATGCAAAATAAAAGTTAGA 60

Db 1314 AAGCTTTAAATTTGGAAAAATTTAATAATATATGATTTAAATGCAAAATAAAAGTTAGA 1372

OY	6	TTGAGAGTTTTTTAAATTTTTTATATAAAAATATTTTTGCATGAATGTTTTAGA	120
Db	1374	TGAAGTTTTTTTTAAATTTTTTATATAAAAATATTTTTGCATGAATGTTTTAGA	1433
OY	121	TAAATTTTGCATPAACCAAGTATTAATTTTCGCCACGAATGAGTCTACGAGATGTC	180
Db	1434	TAAATTTTGCATPAACCAAGTATTAATTTTCGCCACGAATGAGTCTACGAGAGATGTC	1493
OY	181	CTGCTTTAAACCTCGTAAAGTTTTGGTCTTACCCAAACCAATPCCCAAAAGSTAACG	240
Db	1494	CTGCTTTAAACCTCGTAAAGTTTTGGTCTTACCCAAACCAATPCCCAAAAGSTAACG	1553
OY	241	ACCATACCGGTAAATATATTTCTPAACGGTTTAAAGTTTACATTAATCAATTTACTAATC	300
Db	1554	ACCATACCGGTAAATATATTTCTPAACGGTTTAAAGTTTACATTAATCAATTTACTAATC	1613
OY	301	CGCGTGTAATTAAGTTTATGAATGTGTTATTTTGTAGGTCACTGAAATTTATTAAT	360
Db	1614	CGCGTGTAATTAAGTTTATGAATGTGTTATTTTGTAGGTCACTGAAATTTATTAAT	1673
OY	361	TTTTTACTACTGTTTTCTTTTTTGGGTCAACTAGTACTTTTTCTTTGCATCAACAA	420
Db	1674	TTTTTACTACTGTTTTCTTTTTTGGGTCAACTAGTACTTTTTCTTTGCATCAACAA	1733
OY	421	AATATTGTAGACAGAGTGTCCATATATPAGATGTGAAATGAATGAATATTAGTATA	480
Db	1734	AATATTGTAGACAGAGTGTCCATATATPAGATGTGAAATGAATGAATATTAGTATA	1793
OY	481	AATTAATATAGAAAGTGAACAAAAAAATTAAGTGGAAAGAGTAATCTGAAAGAAAGAGCT	540
Db	1794	AATTAATATAGAAAGTGAACAAAAAAATTAAGTGGAAAGAGTAATCTGAAAGAAAGAGCT	1853
OY	541	TATTGGCTTGGCAAGATTCCATTAAGTTTATTTTCACCAAAAGGAAAGATGACTGGCCTT	600
Db	1854	TATTGGCTTGGCAAGATTCCATTAAGTTTATTTTCACCAAAAGGAAAGATGACTGGCCTT	1913
OY	601	GCTCTCTCTCTCTTTTAAATGGTAAACGTGAAGATGAGAGACTCAATTTCTAGTTCA	660
Db	1914	GCTCTCTCTCTCTTTTAAATGGTAAACGTGAAGATGAGAGACTCAATTTCTAGTTCA	1973
OY	661	TCAAAACATATTGGCGGTATATATTCTGTGAGTCACTTGAATAC 704	
Db	1974	TCAAAACATATTGGCGGTATATATTCTGTGAGTCACTTGAATAC 2017	

```

RESULT 3
US-10-473-126-386/C
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473.126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

```

Query Match	9.1%	Score 64.2	DB 21	Length 8056
Best Local Similarity	45.7%	Pred. No. 0.059		
Matches 260	Conservative	0	Mismatches 308	Indels 1
				Gaps 1

QY 5 TTTTAAATTTGGAAAATTTATATATATGTAATTAATGCGCAATATAAGTAGATGAG 64  
 Db 7154 TTAACAAATTATATATATATATTTTATTTAAATTTTAAAAAAATTTAAATTAATAAT 7095







Query Match	8.6%	Score 60.6;	DB 18;	Length 11422;
Best Local Similarity	48.0%;	Pred. No. 0.31;		
Matches 272;	Conservative 0;	Mismatches 284;	Indels 11;	Gaps 3;

Db 7736 TATAGATTAAATTTTATTATT 7762

Query Match	8.6%	Score 60.4;	DB 20;	Length 560;
Best Local Similarity	50.7%	Pred. No. 0.12;		
Matches 145; Conservative	0;	Mismatches 141;	Indels 0;	Gaps 0;

```

491 GAAAGTGACCAAAAAAAAAATTAGTCGAAAAAGGTAACTCGAAGAGAAA 538
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 86

```

; PRIOR FILING DATE: 2000-05-25

```
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60960
; LENGTH: 626
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 48, 49, 50, 117, 135, 136, 137, 166, 168, 189, 194, 197,
; LOCATION: 227, 251, 253, 277, 278, 279, 281, 291, 323, 324, 327, 328,
; LOCATION: 329, 331, 336, 337, 338, 339, 371, 372, 376, 377, 379, 381,
; LOCATION: 382, 394, 403, 404, 405, 406, 408, 413, 429, 431, 437
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 447, 451, 453, 455, 456, 461, 469, 484, 487, 493, 499, 509,
; LOCATION: 515, 529, 538, 540, 541, 542, 545, 558, 559, 561, 564, 565,
; LOCATION: 581, 582, 584, 585, 589, 592, 598, 599, 601, 606, 609, 617,
; LOCATION: 618
; OTHER INFORMATION: n = A,T,C or G
; US-10-357-930-60960

Query Match      8.5%; Score 60.2; DB 21; Length 626;
Best Local Similarity 48.0%; Pred. No. 0.14;
Matches 134; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 258 TATTCACCGGTTATATAGTTTACATTAATCTTACTTAATCCGGGTAAATTAAGTT 317
DB 330 TNNNCNCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCCGCCCNCCNNAATTT 271
QY 318 TATGAAATGCGTTATTTGTAGTCACGTGAAATTTATTAATTTTGTAGTCTGTTT 377
DB 270 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 211
QY 378 TCTTTTGGGTCACTAGTACTTTTCTTTCCTTGACATCAAAATATTTAGACGAGT 437
DB 210 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 151
QY 438 GGTCCATATATAGATGTAATGAAATGAATATTTAGTAAATTAATATAGAAAGTG 497
DB 150 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 91
QY 498 AACCAAAAAAATTAGTGAAGGTAAGTGAAGGAAAGAAA 536
DB 90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 52

RESULT 11
US-10-021-323-1716
; Sequence 1716, Application US/10021323
; Publication No. US2004012340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, J111
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 1716
; LENGTH: 499
; TYPE: DNA
```

```
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3825-005-Q1-N6-D10
; US-10-021-323-1716

Query Match      8.5%; Score 60; DB 20; Length 499;
Best Local Similarity 50.7%; Pred. No. 0.14;
Matches 144; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 305 TGTAAATTAAGTTTATGAAATGTGTTATTTTGTAGTCACGTGAATTTATTAATTTT 364
DB 34 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 93
QY 365 TAGTACTGTTTCTTTTCTTTTGGGTCAAGTACTTTTCTTCTTGACATCAAAAT 424
DB 94 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 153
QY 425 ATTGAGACGAGTGTCCATATATAGATGTAATGAAATGAATATTTAGTAAATATA 484
DB 154 AAAAAAGGGGGTAAAAAAATTTTAAAAAGAAAAAAAGAGGGGGGTAAAAA 213
QY 485 AATATAGAAAGTGAAACAAAAAATTAAGTGAAGGTAAGTGAAGGAAAGGCTTAT 544
DB 214 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGGGAAAAAAA 273
QY 545 GCGTTGCAAGATTCATTAAGTTTATTTCCACCAAAAGGAAAGA 588
DB 274 AAAATGGGAAAAAAGGGGTTTAAAAAAGGGGGGGGGGGGGGGGGGGGGGGGGGG 317

RESULT 12
US-10-473-126-240
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Biogenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hemopoietic cell
; TITLE OF INVENTION: Proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-473-126-240

Query Match      8.4%; Score 59.4; DB 21; Length 8056;
Best Local Similarity 44.7%; Pred. No. 0.46;
Matches 272; Conservative 0; Mismatches 336; Indels 1; Gaps 1;

QY 11 ATTGGGAAAAATTAATAATATATGTAATTAATGCAATTAAGTTAGATGAGTTT 70
DB 2109 ATTATATATATATTAATTAATTTTATTTATTTATTAATAAATAAATAAATAA 2168
QY 71 TTTAATTTTATTAATAAATAATTTTTCGATGAAATTTGTTTAAAGATTAATTTTG 130
DB 2169 ATTATATATTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2228
QY 131 ACATTAACAAGATTAATTTTCGCAAGATTAAGTCAAGAGAGTCTGCTTTAA 190
DB 2229 AATTTTAATTAATAAATTTATTTTAAATTAATTAATTAATAAATAAATAA 2288
QY 191 ACCTGTAAGTTTGTCTTACCAACCAATACCCCAAAAGTAAAGACATACCGG 250
DB 2289 AATTTATTTATTTTAAATAAATAATTTATTTATTTATTTATTTATTTATTT 2348
QY 251 TTAATATATTTCAACCGGTTATTAAGTTTACATTAATCATTTACTAATCCGCTAAT 310
DB 2349 TTTAATATATTTTATTTTATTAATAATTTATTAATTTTATTTTATTTTATTTT 2408
```

[illegible]

```

RESULT 13
US-10-723-860-7676/c
: Sequence 7676, Application US/10723860
: Publication No. US20040253606A1
: GENERAL INFORMATION:
: APPLICANT: Aziz, Natsasha
: APPLICANT: Ginsburg, Wendy M.
: APPLICANT: Zlotnick, Albert
: TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
: TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
: FILE REFERENCE: 05882.0193.NPUS01
: CURRENT FILING DATE: 2003-11-26
: PRIOR APPLICATION NUMBER: 60/429,739
: PRIOR FILING DATE: 2002-11-26
: NUMBER OF SEQ. ID NOS: 8193
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 7676
: LENGTH: 6816
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1919)..(1989)
: OTHER INFORMATION: n is a, c, g, or t
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (3223)..(3247)
: OTHER INFORMATION: n is a, c, g, or t
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (6189)..(6189)
: OTHER INFORMATION: n is a, c, g, or t
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (6205)..(6209)
: OTHER INFORMATION: n is a, c, g, or t
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (6215)..(6221)
: OTHER INFORMATION: n is a, c, g, or t
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (6223)..(6230)
: OTHER INFORMATION: n is a, c, g, or t
: FEATURE:
: NAME/KEY: misc_feature

```

LOCATION: (6234)..(6239)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6241)..(6244)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6246)..(6248)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6250)..(6250)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6261)..(6261)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6276)..(6276)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6276)..(6281)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6291)..(6300)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6307)..(6307)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6305)..(6305)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6322)..(6322)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6324)..(6327)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6338)..(6339)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6341)..(6342)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6351)..(6351)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6386)..(6386)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6388)..(6388)	n	is	a,	c,	g,	or	t
OTHER INFORMATION: n	is	a,	c,	g,	or	t	
FEATURE:							
NAME/KEY: misc_feature							
LOCATION: (6413)..(6414)	n	is	a,	c,	g,	or	t

```

? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (6666) .. (66660)
? OTHER INFORMATION: n is a, c, g, or t
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (6665) .. (66663)
? OTHER INFORMATION: n is a, c, g, or t
?
? FEATURE:

```

Qy	5	TTTTAAATTTGGGAAAATTTAAATATATATAGTTTAAATGGCAAAATAAAGGTAGATGAC	64
Db	6653	TTTATATANNNAATTAATAAAAAAAAAAATATATANNNAAAAAAAAAAAAAAAAAAAAAAT	6594
Qy	65	AGTTTTTTAAATTTTTTATATATAAATATTTTTCATGAATTCGTTTTTAAAGATAAA	124
Db	6593	TTTTTTATATAATTTTTTTTAAAAAANNTTTTTTTTTTAAAAAATATTTTAAATANN	6534
Qy	125	ATTTTGCATTAACCAATATATTTTCCGCGAAGATGAGTACGAGAGATGTCCTGT	184
Db	6533	TTTTTATTTTTAAATTTTATTTTAAAAATAAAAATATATTAATAAANNTATTTTTTT	6474
Qy	185	CTTTAAACCTCGTAAAGTTTGGCTTAACCCAAACCAATACCCACAAAGGTAAACGACA	244
Db	6473	AAAAAANNTNAAAAATTTTTTTTTTATTTTTT-----TTNCNNAAAAAATATAATATAA	6419
Qy	245	TACCGGTAATATAATCTCAACCGGTTATTAAGTTTACATTAATCATTTACTAATCCCGG	304
Db	6418	NTNANNTAAAAAATTTTTTTTTTAAAAATAANNTTTTTTATATAAAAAAATTTTTTT-----	6364
Qy	305	TGTATATTAAGTTTATGAATANGGTAATTTTGGAGGCAAGTAAATTTATTAATTTTT	364
Db	6363	--TAAAAAANNTATAAATTTNNNTTTTTTTTATTTNNNNNTTATATTTATTTTTTNT	6306
Qy	365	TAGTACTGTGTTTTCTTTTTGGGCTCAACTAGTTACTTTTTCTTTGACATCAAAAT	424
Db	6305	NTTTNNNNNNNNNTTTTTTTTTNNNNNTTTTTTTTTTTTTNTNNNNNTTNTNTNT	6246

Query Match	8.4%;	Score 59.2;	DB 22;	Length 50000;
Best Local Similarity	54.8%;	Pred. No. 0.94;		
Matches 138;	Conservative	0;	Mismatches 113;	Indels 1;
			Gaps	1;

QY 260 TTCTAACCGTTTATAGTTTACATAAATCATTTACTAATCCGGTGAATTAAGTTT 319  
DB 49095 TTATTTACTTATTTTATATTTTATTCATCATTTTATATATTAATTAATTTTCAATATAT 49154  
QY 320 TGAATGTGGTTTATTTTGTAGCGTCAAGTGAATTTATTAATTTTGTAGTCTGTTTTC 379  
DB 49155 TATATATTTTATTTTATTTTATTAAGATATTTTATTAATTTTGTATATTTATTTAT 49214  
QY 380 TTTTGTGGTCACTAGTACTTTTCTTGTACATCAAAATTAATTTGTAGACGAGTGG 439  
DB 49215 AAATTTTATTAATATATATATATTTGTTTAATTTTAATCAATATATTTTAAACATGCCA 49274  
QY 440 TCCATATATAGATGTGAATGAATGAATATTTGAGTAAATTAATATAGAAAGTGA 499  
DB 49275 AAAATGAATATATAGTAAATTTTATTTATA-GATTTTTCATTAATAATATATATTTGTT 49333  
QY 500 CAAAAAAATTA 511  
DB 49334 CTAAATATATTA 49345

## RESULT 15

US-10-021-323-16830  
; Sequence 16830, Application US/10021323  
; Publication No. US20040123340A1  
; GENERAL INFORMATION:  
; APPLICANT: Delkman, Jill  
; APPLICANT: Feng, Paul C.C.  
; APPLICANT: Fincher, Karen L.  
; APPLICANT: Ziegler, Todd E.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(52274)B  
; CURRENT APPLICATION NUMBER: US/10/021,323  
; PRIOR FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255, 619  
; PRIOR FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 17880  
; SEQ ID NO 16830  
; LENGTH: 469  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3829-024-Q6-N6-E4  
US-10-021-323-16830

Query Match 8.4%; Score 59; DB 20; Length 469;

Best Local Similarity 51.3%; Pred. No. 0.21;

Matches 137; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 270 TTATTAAGTTTACATTAATCATTTACTAATCCGGTGAATTAAGTTTATGAAATGTGG 329  
DB 11 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 70  
QY 330 TTATTTGTAGTCAAGTGAATTTATTAATTTTGTAGTCTGTTTCTTTTGTGGT 389  
DB 71 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 130  
QY 390 TCACTAGTACTTTTCTTGTACATCAAAATATTTGTAGACGAGTGTCCATATATA 449  
DB 131 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 190  
QY 450 GATGTGAATGAATGAATATTTAGTATTAATAATTAAGTGAACAAAAAAT 509  
DB 191 AAAAAAAAAAAAAAAAAATTTTATTAATAAAAAAAAAAAAAAAAAAAAAA 250  
QY 510 TAGTGAAAAAGTAACTGAAAAAGAAA 536  
DB 251 TTTAATAAAAAAAAAAAAAAAAAAAAA 277

Search completed: November 17, 2005, 18:03:54  
Job time : 855 secs

*This Page Blank (uspto)*